Districting for Parcel Delivery Services { A Two-Stage Solution Approach and a Real-World Case Study

Citation for published version:
Bendera, M, Kalcsics, J & Meyer, A 2020, 'Districting for Parcel Delivery Services { A Two-Stage Solution Approach and a Real-World Case Study', *Omega*, vol. 96, 102283.
https://doi.org/10.1016/j.omega.2020.102283

Digital Object Identifier (DOI):
10.1016/j.omega.2020.102283

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
Omega

Districting for Parcel Delivery Services – A Two-Stage Solution Approach and a Real-World Case Study

Matthias Bender[a], Jörg Kalcsics[b], and Anne Meyer[c]

[a]FZI Research Center for Information Technology, 76131 Karlsruhe, Germany
[b]School of Mathematics, University of Edinburgh, Edinburgh, EH9 3FD, United Kingdom
[c]Faculty of Mechanical Engineering, TU Dortmund University, 44227 Dortmund, Germany

May 27, 2020

Abstract

This paper studies a real-world problem arising in the context of parcel delivery. Given a heterogeneous set of resources, i.e., different drivers and different vehicles, the problem for each day consists of assigning a driver and a vehicle to each customer requiring service. Two conflicting aspects must be taken into account. On the one hand, service consistency is desirable, meaning that a customer should always be served by the same driver. On the other hand, daily demand fluctuations and tight resource constraints prohibit fixed resource assignments. With the aim of finding a reasonable compromise between these aspects, we propose a novel two-stage districting approach, which establishes delivery districts in the first stage and adapts them to the daily demand realizations in the second stage. For the first stage problem we propose three models that differ in the level of detail of their input data, their expected compliance with service consistency and the driver's contractual working times, and their computational effort. Our two-stage approach merges the two dominant approaches in the literature, which either determine a priori routes and then adapt them on a daily basis, or derive fixed service regions for drivers. We present a case study based on a real-world data set. The results highlight the differences between the three first stage models and show that only few adaptations of the districts are necessary in the second stage to achieve feasible daily delivery tours along with a very good workload balance for drivers. We also analyze the effects of a homogeneous vs a heterogeneous fleet, of full time drivers vs full and part time drivers, and of the location of the depot and the length of the planning horizon.

1 Introduction

Parcel delivery companies, such as UPS, FedEx, and DHL, deliver myriads of packages to customers every working day. Solely in Germany 3.35 billions packages were shipped in 2017 with expected growth rates of around 5%[1]. Although recently new concepts, such as pick-up shops or pick-up stations, have arisen, a huge part of the packages is delivered by drivers directly to the homes of customers. In these cases, parcels for a service region are sorted in a central depot and then delivered to customers by tours starting and ending at that depot. The tasks are performed by a heterogeneous fleet of vehicles with different capacities and a heterogeneous crew of drivers with different contractual working times.

In this paper, we study the planning problem arising at clients of our project partner PTV Group (PTV), a world leading provider of optimization solutions for territory design and vehicle routing problems: On a tactical level, i.e., for the next weeks or months, we have to determine the crew of drivers and the fleet of vehicles; hereby, drivers may have different contractual working times and vehicles may have different sizes. On an operational level, for each working day we have to assign vehicles and drivers to customers requiring service on this day—at UPS this planning step is referred to as van assignment problem Holland et al. [2017]. This must be done in a way that the capacity of the respective vehicles is not exceeded and overtime for drivers is avoided. Besides aiming for short delivery tours, important goals are to balance the workload between drivers, to keep the delivery district of each driver as compact as possible, and to provide driver consistency over time when assigning customers to drivers.

The reasons for these additional goals are operational: If we manage to balance the workload of the drivers relative to their contractual working times, then also the overtime (if necessary) is fairly balanced over the crew of drivers. Geographical compact districts grant a high degree of flexibility to drivers with respect to the sequence in which they visit customers without overly increasing the distance traveled. This can be highly advantageous if additional constraints restrict the sequence of the tour, such as time windows, or if the need for a second visit to a commercial customer to pick up parcels later on only becomes known during the tour [also a common situation at UPS, see Holland et al., 2017]. Furthermore, in compact districts the impact of traffic congestion is smaller due to on average shorter edges comprising a tour. This is especially true for urban areas. Aiming for driver consistency has two reasons: First, drivers become familiar with their delivery districts, which increases their efficiency to provide service to the customers Smilowitz et al. [2013], Zhong et al. [2007]. Second, if always the same driver visits a certain customer, then this establishes a personal connection between drivers and customers Groër et al. [2009] and increases customer satisfaction Jarrah and Bard [2012]. More generally, this form of consistency is called person-oriented consistency and is common also in other applications, e.g. in home health care provision. For more details on this and other types of consistency, see Kovacs et al. [2014].

There exist two main approaches in the literature to solve these problems: generate a priori routes and make daily adaptations, and assign drivers to fixed regions Kovacs et al. [2014]. Starting with the first, the problem of assigning customers to vehicles and drivers and determining a visit sequence resulting in short tours can be treated as a classical vehicle routing problem (VRP) solved individually at the beginning of each day. However, with a high fluctuation in demand and tight resources, as it is common in parcel delivery, routes will often be inconsistent Kovacs et al. [2014]. Thus, the classic VRP needs to be extended considerably to additionally consider driver

[1]KEP Studie 2018, German Association of CEP service provider (Bundesverband der Kurier-Express-Post-Dienste e.V.) `https://www.biek.de/download.html?getfile=1928` (26.10.2018)

consistency, guarantee a certain degree of compactness, build balanced tours with respect to the workload, and consider a heterogeneous fleet of vehicles and drivers. These extensions render VRP models extremely difficult to solve in their entirety [for recently published, thorough discussions about relevant extensions and variants see Vidal et al., 2019, Rossit et al., 2018, Matl et al., 2017]. Additionally, even UPS, a company known for its high degree of automation and digitization within its package centers, reports that the exact addresses that a driver has to visit on a given day is only known after the vehicles have been loaded with parcels Holland et al. [2017].

In contrast to VRP approaches, districting approaches naturally consider customers on an aggregated level (basic areas), e.g. on the level of streets instead of exact addresses. The goal of districting approaches is to group these basic areas into geographically compact and, in terms of workload, balanced delivery districts. Districting implicitly ensures long-term service consistency, if the delivery districts are kept fairly stable over a long period of time, and the same driver, or the same team of drivers, is responsible to serve all customers in a district during this time period [for a general introduction to districting we refer the reader to Kalcsics, 2015]. Based on these observations, districting approaches are an alternative to VRP approaches. However, different challenges have to be addressed now: As customer demand is fluctuating, it is not possible to strictly hold on to the delivery districts determined on the tactical level. Delivery districts must be adapted on a day-to-day basis, just as it is necessary in a vehicle routing approach. Furthermore, the workload of a district can only be approximated, since the workload of a driver depends on the driving time and, hence, on the concrete tour, which is not determined in a districting approach.

To the best of our knowledge, no approach in the literature addresses the problem described above in its entirety. In this contribution, we show the potential of districting approaches to assign customers to vehicles for parcel deliveries. As we also allow for daily adaptations, our approach merges the two most common ones identified in Kovacs et al. [2014]. To account for demand fluctuations, we treat the problem as a two-stage problem, as it is common in practice Wong [2008], with the two stages corresponding to different planning levels. On both planning levels, we are faced with conflicting objectives between which a reasonable compromise must be found: (1) On a tactical planning level, we must subdivide the service region into an adequate number of delivery districts, and we have to assign resources to each district. Hereby, we have to find a reasonable trade-off between the number of districts, and, hence, the number of required resources, and the expected workload of the districts. (2) On an operational level, i.e., on a day-to-day basis, we have to adapt districts to the concrete demand realization of a day, while preserving the resource assignments made at the tactical level. Here, it is important to find a good compromise between service consistency and working time related objectives, such as compliance with the drivers' contractual working times and workload balance between the drivers.

The main contributions of this paper are as follows:

- We propose a novel two-stage solution approach for the assignment of drivers and vehicles to customers that relies on districting instead of vehicle routing techniques and allows for daily adaptations, thereby merging the two most common solution approaches.

- For the tactical planning level, we propose a districting approach that involves the determination of the number of districts and the assignment of heterogeneous resources. This combination has, to the best of our knowledge, not been considered in the districting literature before.

- We present three integer programming (IP) models for the tactical planning problem, which differ in the level of detail of their input data and in their expected compliance with the drivers' contractual working times. Moreover, we present a heuristic solution procedure for the tactical problem.
- For the operational level, we propose a mixed integer programming (MIP) model that adapts the tactical districting solution to the concrete demand realization of a day.
- We perform an extensive case study based on real-world data, to test the effectiveness of our districting based approach. In particular, we analyze the feasibility of using districting approaches for the problem at hand and the suitability of the three tactical planning models, and we investigate the trade-off between compliance with the drivers' contractual working times and service consistency.

The paper is organized as follows. We review related literature in Section 2, and describe the problem under study in detail in Section 3. We present our two-stage mathematical model in Section 4, and the corresponding solution approach in Section 5. The measures used to evaluate solutions are explained in Section 6. Section 7 contains the results of the case study. We provide some concluding remarks in Section 8.

2 Related Work

We restrict our literature review to districting problems for vehicle routing applications in which demand uncertainty plays an important role. In such a setting, districting approaches that cluster the customers on a tactical level and decide on a routing on a daily basis, are particularly attractive as they do not only implicitly provide service consistency, but also entail administrative convenience and facilitate daily route planning Wong and Beasley [1984]. Assigning drivers to fixed regions was characterized in Kovacs et al. [2014] as one of the two main approaches to ensure person-oriented consistency, among other things, in the presence of fluctuating demands and tight resource constraints. For more details on the other approaches, we refer the reader to their paper.

We give an overview of models dedicated to designing or adapting delivery districts in presence of uncertain demand in Table 1. We classify the models according to the planning criteria that differentiate our application from classical districting models: Both the tactical planning (TP) of districts and the operational planning / adaptation (OP) based on the concrete demand realization needs to be considered. At the tactical level the number of necessary districts needs to be determined (DND), and heterogeneous vehicles and drivers need to be assigned to districts (RA).

The authors of the first row of Table 1 propose heuristics for the tactical design of delivery districts based on uncertain demand. However, none of them addresses the determination of the number of districts, the assignment of inhomogeneous resources, or the operational decision stage.

Concerning the second set of works, the author of Haughton [2008] investigates in a simulation study the impact of fixed districts on routing efficiency by comparing a districting approach with

Table 1: Overview of planning criteria considered in the related literature. The works are categorized according to the combination of considered planning stages and planning criteria.

TP	OP	DND	RA	Work
✓	–	–	–	Wong and Beasley [1984], Galvão et al. [2006], Ouyang [2007], González-Ramírez et al. [2011], Carlsson [2012], Carlsson and Delage [2013]
✓	✓	–	–	Zhong et al. [2007], Haughton [2008], Schneider et al. [2015]
✓	–	✓	–	Haugland et al. [2007], Bard and Jarrah [2009], Lei et al. [2012, 2016]
✓	✓	✓	–	Daganzo and Erera [1999], Erera [2000]
✓	–	✓	(✓)	Novaes and Graciolli [1999][1], Novaes et al. [2000][1], Jarrah and Bard [2012][2]
–	✓	–	–	Janssens et al. [2015], Holland et al. [2017]
✓	✓	✓	✓	This paper

[1] Although different vehicle types can be considered, solutions always consist of a homogeneous fleet.
[2] Although different vehicle types are supported in principle, this design feature is not exploited.

a daily route optimization from scratch. As before, the number of districts was given beforehand, and the resources were assumed to be homogeneous.

Most closely related to ours are the applications treated in the works of Zhong et al. [2007] and Schneider et al. [2015]: Both approaches try to find a reasonable trade-off between consistency and routing flexibility in the context of parcel delivery. The authors of Zhong et al. [2007] introduce the concept of "core areas", which correspond to partial districts that are served by the same driver every day. Basic areas not assigned to a "core area" are free to be assigned to any core area, or even to extra drivers, on a day-to-day basis. To foster consistency during the operational planning, the authors consider driver learning by assuming that the average time spent in a basic area decreases with an increasing number of continued visits to that area. The authors propose a tabu search for the strategic core area design, and a method based on a parallel insertion heuristic for the operational planning stage.

The authors of Schneider et al. [2015] also propose a two-stage solution approach. Following the heuristic of Wong and Beasley [1984] for the first stage, they design partial districts based on vehicle routing solutions for historical sample days. Those partial districts are similar to "core areas" and are built by assigning a certain percentage of customers in a greedy fashion to selected seed customers, preferring joint assignments of customers often served on the same historical tours. In the second stage, they solve a VRP with time windows for a concrete demand realization of a day, using a tabu search heuristic. This heuristic considers the district assignments of the first stage and allows for a certain number of customers to be reassigned. As the customers not assigned to a partial district may end up in a different tour every day, this approach considers driver consistency only to a certain extent in the second stage.

While the works of Zhong et al. [2007] and Schneider et al. [2015] address both decision stages, they assume that the number of districts is given in advance and that the fleet of drivers and vehicles is homogeneous during the tactical planning.

In the approaches of the third row of Table 1, the number of districts is a result of the tactical planning stage: The authors of Haugland et al. [2007] propose a tabu search heuristic that minimizes the expected total travel cost, considering an upper bound on the travel cost within each district. The authors of Bard and Jarrah [2009] propose a set-partitioning formulation and solve it using heuristically generated clusters. The goal of the algorithm is to reach the minimum number of districts, such that the expected workload and weight in each district can be handled by the capacity of a single vehicle and within the drivers' working time. The authors of Lei et al. [2012] and Lei et al. [2016] propose two-stage stochastic programs to design delivery districts comprising regular and stochastic customers in a single-period and a multi-period setting, respectively. The objectives are, beyond others, to minimize the number of districts and the expected routing costs. For the single-period case, they propose a large neighborhood search, and for the multi-period case, a multi-objective evolutionary algorithm.

In contrast to the approaches mentioned before, Daganzo and Erera [1999] and Erera [2000] not only determine the number of districts and their design, but also address the operational planning stage. The authors propose different strategies to deal with actual demand realizations on the day-to-day level, including, for example, "sweeper tours" that contain all unserved customers of the initial tours, as well as more sophisticated schemes that involve the dynamic coordination of vehicles in real time. However, the district design is not based on exact, but continuous approximation models, and the proposed approaches assume identical capacities in terms of load and available driving time.

With the objective of minimizing total daily transportation cost, Novaes and Graciolli [1999] and Novaes et al. [2000] determine the number and the design of delivery districts. They propose to partition the service region into districts using a ring-radial pattern, and use approximation formulas to compute expected tour lengths. Although their models can take different vehicle capacities and operating costs into account, their approaches can only obtain solutions for a homogeneous fleet of vehicles. In a follow-up paper to Bard and Jarrah [2009], the authors of Jarrah and Bard [2012] propose a column-generation approach that uses heuristics to generate clusters in the pricing phase. Their approach supports, in principle, the construction of clusters of various capacities, corresponding to the capacities of different available vehicles. But the authors state that they do not exploit this design feature and consider only test data with a homogeneous fleet. Neither the approaches of Novaes and Graciolli [1999], Novaes et al. [2000], nor the approach of Jarrah and Bard [2012] considers different working times for drivers or the operational planning stage.

The authors of Janssens et al. [2015] address the situation in which a tactical districting plan, which assigns basic areas (referred to as "microzones") to vehicles, is given. In order to balance workloads and to achieve feasible tours, at the operational planning stage the basic areas are reassigned corresponding to the actual demand of a day. The authors propose a multi-neighborhood tabu search heuristic that minimizes the total transportation cost, the deviation from the tactical plan, and workload imbalances. However, the authors assume that vehicles are not capacity-constrained. In Holland et al. [2017] the planning system of UPS is described. On the tactical level, the system determines one base tour per depot. Once a concrete demand realization becomes known on the operational level, each base tour is cut into a daily varying number of vehicle tours depending on their capacities. Afterwards, the system determines the sequence of each vehicle, aiming for consistency compared to the sequence of the base tour and considering further requirements for guaranteeing operationally sensible tours (e.g. no zigzagging on busy streets). However, the authors do not describe the approach for determining the base tour. Moreover, the vehicle assignment procedure does not explicitly model driver consistency and is, according to the authors, based on the "existing dispatching practice" and subject of ongoing improvement.

Since none of the described approaches addresses all features of our problem, we propose a new two-stage approach following the two-stage nature of our problem: For the tactical planning stage, we introduce a solution approach that combines the joint determination of the design and the number of districts, with the assignment of heterogeneous vehicles and drivers. For the operational planning stage, we propose a model that reassigns basic areas based on the concrete demand realization of a day and strives for consistency with respect to the driver and vehicle assignments from the previous stage. During both stages, we consider the possibility to return to the depot for reloading, a feature addressed in none of the approaches introduced before.

3 Problem Description

In this section, we describe in detail the planning problem along with the relevant input data, and we introduce the notation that is necessary for the models proposed in Section 4. In the appendix, we include a summary of the notation in Table 5.

3.1 Tactical Design

In the tactical planning problem, the task is to partition the set of *basic areas* $B = \{1, ..., |B|\}$ belonging to a service region into *delivery districts*, and to assign a *driver type* $d \in D = \{1, ..., |D|\}$ and a *vehicle type* $v \in V = \{1, ..., |V|\}$ to each district.

Each basic area $b \in B$ represents a geographical area in the service region, e.g. a zip code area. For basic areas $b, i \in B$, $c_{bi} \in \mathbb{R}^+$ denotes the distance between b and i. We denote by NG the neighborhood graph of all basic areas. In this graph, each basic area coincides with a vertex and two basic areas are connected by an edge, iff they are geographically adjacent to each other. We will explain in Section 7.1 how we generate this graph and determine geographical adjacency. Moreover, $A_b \subseteq B$ defines the set of all basic areas that are adjacent to basic area $b \in B$ in NG.

Driver types distinguish themselves by the drivers' contractual daily working times in relation to a full-time driver. The contractual working time of a full-time driver is given by $t^{max} \in \mathbb{R}^+$, and the relative contractual working time of driver type $d \in D$, expressed as the percentage of t^{max}, is denoted as $r_d \in (0, 100]$. The number of available drivers of type $d \in D$ is given by $M_d \in \mathbb{N}^+$. For every vehicle type $v \in V$, its capacity is given by $C_v \in \mathbb{R}^+$, while $N_v \in \mathbb{N}^+$ denotes the number of available vehicles of type $v \in V$. For both resource types, we assume that they can be totally ordered, i.e., we know for each pair of resource types which type is preferred from an economic or operational point of view and, thus, should be used with higher priority. As an example, permanent employees might be preferred over contract workers, or own vehicles might be preferred over vehicles that are available for leasing. Formally spoken, driver type d_1 is preferred over driver type d_2 if and only if $d_1 < d_2$. The same holds for two types of vehicles.

Demand data is given for a set $T = \{1, \ldots, |T|\}$ of sample days in the form of *customer orders* $O = \{1, ..., |O|\}$ comprising all parcels with the same delivery address and the same delivery day. For each customer order $o \in O$ we know the total weight of the parcel(s) $l_o \in \mathbb{R}^+$, the service time $s_o \in \mathbb{R}^+$, the delivery day $\tau_o \in T$, the delivery location, and, thereby, the basic area $b_o \in B$ that contains the customer. Ideally, this data is available as forecast data, i.e., it represents expected future demand.

Each delivery tour starts and ends at a given depot. We denote by $t_i^{depot} \in \mathbb{R}^+$ the time required to travel between the depot and basic area $i \in B$. If the capacity of a vehicle does not suffice to serve all customers that require service by the vehicle, the vehicle must return to the depot, must be reloaded, and another delivery tour has to be made. In this case, a vehicle reloading time, $t^{reload} \in \mathbb{R}^+$, is incurred. Following the convention in the VRP literature, we refer to tours with at least one return to the depot for reloading as *multi-trip* tours.

When we design the delivery districts and assign the resources based on the demand of the sample days T, we must not exceed the maximum number of drivers M_d of each driver type $d \in D$ and the maximum number of vehicles N_v of each vehicle type $v \in V$. Moreover, we must respect the preference of driver types, i.e., a driver of type $d \in D$ may only be used, if for each driver type $d' \in D$ with $d' < d$, all $M_{d'}$ available drivers are also used; analogously for vehicle types. The districts are supposed to be contiguous and geographically compact, as this facilitates the construction of short delivery tours on a daily basis. The number of delivery districts is not given in advance, but it is part of the problem to determine an adequate number. On the one hand, the number of districts has to be sufficiently large such that the size of the districts allows each driver to serve the customers in his district without working overtime. On the other hand, establishing more districts than required to meet the demand results in a low utilization of the assigned resources and, hence, is inefficient. Consequently, it is important to find a good trade-off between compliance with the drivers' contractual working times and resource efficiency.

3.2 Operational Adaptation

Once the districts have been designed on the tactical level, the task is now to adapt them on a day-to-day basis to a concrete demand realization of orders $O^{op} = \{1, ..., |O^{op}|\}$ for a single operational day, where $o \in O^{op}$ is defined as above. For this purpose, we are allowed to modify the assignments of basic areas to delivery districts. However, we must not change the assignments of driver types to delivery districts, as this would eliminate consistency. Furthermore, the vehicle type that is assigned in the tactical decision must be preserved. This means that we have to partition the set of basic areas B into geographically compact and contiguous delivery districts, while respecting the decisions that were made at the tactical level with regard to the assignments of driver and vehicle types. Since the tactical district design is given, this can also be viewed as a reassignment decision.

As in the tactical problem, we are faced with conflicting goals: On the one hand, we want to strive for consistency, which implies that no or only few basic areas should be reassigned. On the other hand, we have to avoid overtime and unbalanced workloads. Hence, we must find a reasonable trade-off between consistency and working time related objectives.

4 Two-Stage Mathematical Model

In this section, we present our two-stage mathematical model. We give an overview of the planning decisions and data requirements for each stage in Figure 1.

1. Tactical districting

- Partition the set of basic areas into an adequate number of delivery districts
- Assign a driver type to each district
- Assign a vehicle type to each district
- Data basis: Demand data for tactical sample days, aggregated over basic areas

Delivery districts, ↓ resource assignments

2. Operational reassignment

- Partially reassign basic areas according to concrete demand realization
- Preserve driver and vehicle assignments of tactical solution
- Data basis: Demand data for operational sample days, aggregated over basic areas

Figure 1: Overview of the two-stage solution approach

4.1 Stage 1: Tactical Districting

In this section, we propose three different IP models for the tactical design problem. On top of the notation presented in Section 3, we start by introducing additional notation that is common to all models. We measure geographical compactness as the sum of the distances between all basic areas that belong to a certain district and the basic area which is selected as the district center. Such a center-based approach to measure compactness is quite common in the literature on districting [see, e.g., Fleischmann and Paraschis, 1988, Ríos-Mercado and López-Pérez, 2013], and can relatively easily be handled by modern general-purpose MIP solvers. We introduce the following decision variables:

$$x_{bi} = \begin{cases} 1 & \text{if basic area } b \in B \text{ is assigned to the delivery district represented by center } i \in B \\ 0 & \text{otherwise} \end{cases}$$

$$y_{di} = \begin{cases} 1 & \text{if driver type } d \in D \text{ is assigned to the delivery district represented by center } i \in B \\ 0 & \text{otherwise} \end{cases}$$

Note that $x_{ii} = 1$ implies that $i \in B$ is selected as a district center.

The decision variables describing the assignments of vehicle types to districts are model dependent, and we will introduce them in the subsequent sections. Additionally, we introduce the following auxiliary variables which are required to incorporate the preference criteria with respect to different driver and vehicle types:

$$e_d = \begin{cases} 1 & \text{if all available drivers of type } d \in D \text{ are assigned to delivery districts} \\ 0 & \text{otherwise} \end{cases}$$

$$f_v = \begin{cases} 1 & \text{if all available vehicles of type } v \in V \text{ are assigned to delivery districts} \\ 0 & \text{otherwise} \end{cases}$$

We denote by $l_b^\tau = \sum_{o \in O, \tau_o = \tau, b_o = b} l_o$ the total weight of parcels to be transported to customers in basic area $b \in B$ on day $\tau \in T$. Furthermore, $w_b^\tau \in \mathbb{R}^+$ states the estimated workload of basic area $b \in B$ on day $\tau \in T$ within the delivery district: It consists of the total service time $s_b^\tau = \sum_{o \in O, \tau_o = \tau, b_o = b} s_o$ and the estimated total travel time within the district that is required to serve all customers in basic area $b \in B$ on day $\tau \in T$. We will explain in Section 7.2 how we estimate the total travel time. Finally, we estimate the travel time between the depot and the delivery district based on the time t_i^{depot} required to travel between the depot and the basic area $i \in B$ that represents the center of the district. Due to vehicle capacity limitations it might be necessary to perform several trips to a delivery district to meet all demands on a given day. Hence, when $n \in N = \{1, ..., |N|\}$ trips are performed to the district represented by basic area $i \in B$, the travel time between the depot and the district plus the time required to reload the vehicle at the depot is given by $t_{ni} = 2 \cdot n \cdot t_i^{depot} + (n-1) \cdot t^{reload}$.

In the following, we present three IP models for the design of delivery districts on the tactical planning level. As Table 2 shows, the models differ in the following two aspects: (1) The models distinguish themselves by the level of detail of their input data for each basic area. (2) The models differ in the way in which workload limits are taken into account for each district.

The level of detail of the input data relates to the estimated workload of each basic area and to the weight that must be transported to each basic area, both of which can be considered either as average values over the $|T|$ tactical planning days (AV) or as individual values for each day (IV).

The workload limits restrict the estimated workload of each district to the interval $[LB, UB]$ with LB, $UB \in \mathbb{R}^+$. The bounds are necessary to prevent the models from creating either very small districts, which result in an inefficient utilization of resources, or very large districts, which lead to considerable overtime. Depending on the modeling variant, these workload limits are either applied to the average daily workload estimation (AW) or to the workload estimation of each individual day (IW).

The characteristics of the three models are summarized in Table 2 and can be described as follows:

- Model AV–AW uses the average daily workloads and weights of the basic areas as input, and applies the workload limits to the average daily workload estimation of the districts. The number of trips to a district depends on the average daily weight that must be transported to the district.

- Model A/IV–AW takes into account the average daily workloads and the day-specific weights of the basic areas. It applies the workload limits to the average daily workload estimation of the districts. As this model considers day-specific weights, the number of trips to a district varies from day to day and depends on the weight that must be transported to the district on each day.

- Model IV–A/IW considers day-specific workloads and day-specific weights of the basic areas. The lower bound is applied to the average daily workload estimation of the districts, whereas the upper bound relates to the individual workload estimation for each day. As in model A/IV–AW, the number of trips to a district varies from day to day depending on the weight to be transported.

Model IV–A/IW is the most conservative of the three models in the sense that we expect it to yield the fewest overtime hours in the operational stage of all models, due to the very restrictive workload upper bound. Model A/IV–AW is expected to be more conservative than model AV–AW, since it takes into account day-specific weights, whereas model AV–AW considers only average weights, and, thus, fluctuations in weight are leveled out. Thus, we expect to obtain more workload peaks in solutions computed with the latter model than in solutions obtained with the former. In Section 7.3, we empirically evaluate if the models behave in the expected way.

4.1.1 Model AV–AW

We introduce the following additional parameters: By $\overline{w}_b = \frac{1}{|T|} \cdot \sum_{\tau \in T} w_b^\tau$ we denote the average daily workload estimation, and by $\bar{l}_b = \frac{1}{|T|} \cdot \sum_{\tau \in T} l_b^\tau$ the average daily weight in basic area $b \in B$. Furthermore, we define binary decision variables:

$$z_{nvi} = \begin{cases} 1 & \text{if vehicle type } v \in V \text{ is assigned to the district represented by basic area } i \in B \\ & \text{and } n \in N \text{ trips to the district are performed} \\ 0 & \text{otherwise} \end{cases}$$

Using this notation, we formulate model AV–AW as follows:

$$\sum_{b \in B} \sum_{i \in B} c_{bi} x_{bi} \rightarrow \min \tag{1}$$

$$\text{s.t.} \quad \sum_{i \in B} x_{bi} = 1 \qquad b \in B \tag{2}$$

Table 2: Overview of the three proposed models

Model	Basic area input data				District workload limits on
	Average **V**alues		**I**ndividual **V**alues		
	Workload	Weight	Workload	Weight	
AV-AW	√	√			$LB \leq$ **A**vg. **W**orkload $\leq UB$
A/IV-AW	√			√	$LB \leq$ **A**vg. **W**orkload $\leq UB$
IV-A/IW			√	√	$LB \leq$ **A**vg. **W**orkload and **I**nd. **W**orkload $\leq UB$

$$x_{bi} \leq x_{ii} \qquad b, i \in B \tag{3}$$

$$\sum_{b \in \bigcup\limits_{b' \in S} (A_{b'} \setminus S)} x_{bi} - \sum_{b \in S} x_{bi} \geq 1 - |S| \qquad i \in B, S \subseteq B \setminus (\{i\} \cup A_i), S \neq \emptyset \tag{4}$$

$$\sum_{d \in D} y_{di} = x_{ii} \qquad i \in B \tag{5}$$

$$\sum_{i \in B} y_{1i} \leq M_1 \tag{6}$$

$$\sum_{i \in B} y_{di} \leq M_d e_{d-1} \qquad d \in D, d > 1 \tag{7}$$

$$M_d e_d \leq \sum_{i \in B} y_{di} \qquad d \in D, d < |D| \tag{8}$$

$$\sum_{n \in N} \sum_{v \in V} z_{nvi} = x_{ii} \qquad i \in B \tag{9}$$

$$\sum_{n \in N} \sum_{i \in B} z_{n1i} \leq N_1 \tag{10}$$

$$\sum_{n \in N} \sum_{i \in B} z_{nvi} \leq N_v f_{v-1} \qquad v \in V, v > 1 \tag{11}$$

$$N_v f_v \leq \sum_{n \in N} \sum_{i \in B} z_{nvi} \qquad v \in V, v < |V| \tag{12}$$

$$\sum_{b \in B} \bar{l}_b x_{bi} \leq \sum_{n \in N} \sum_{v \in V} n C_v z_{nvi} \qquad i \in B \tag{13}$$

$$n z_{nvi} \leq \sum_{b \in B} \frac{\bar{l}_b}{C_v} x_{bi} + (1 - \epsilon) \qquad n \in N, v \in V, i \in B \tag{14}$$

$$LB \sum_{d \in D} \frac{r_d}{100} y_{di} \leq \sum_{b \in B} \overline{w}_b x_{bi} + \sum_{n \in N} \sum_{v \in V} t_{ni} z_{nvi} \qquad i \in B \tag{15}$$

$$UB \sum_{d \in D} \frac{r_d}{100} y_{di} \geq \sum_{b \in B} \overline{w}_b x_{bi} + \sum_{n \in N} \sum_{v \in V} t_{ni} z_{nvi} \qquad i \in B \tag{16}$$

$$e_d \in \{0, 1\} \qquad d \in D, d < |D| \tag{17}$$

$$f_v \in \{0, 1\} \qquad v \in V, v < |V| \tag{18}$$

$$x_{bi} \in \{0, 1\} \qquad b, i \in B \tag{19}$$

$$y_{di} \in \{0, 1\} \qquad d \in D, i \in B \tag{20}$$

$$z_{nvi} \in \{0, 1\} \qquad n \in N, v \in V, i \in B \tag{21}$$

In the objective function (1), we optimize geographical compactness by minimizing the sum of the distances between the district centers and their assigned basic areas. Constraints (2) make sure that each basic area is assigned to a delivery district, and Constraints (3) state that basic areas can only be assigned to delivery districts that are represented by a basic area which is selected as a district center. Constraints (4) were proposed in Drexl and Haase [1999] and ensure the contiguity of the delivery districts. Each of these constraints considers a district center $i \in B$ and a non-empty subset S of basic areas which are not adjacent to the district center. If all basic areas of S are assigned to district center i, then $\sum_{b \in S} x_{bi} = |S|$ and the constraint reduces to $\sum_{b \in \bigcup_{b' \in S} (A_{b'} \setminus S)} x_{bi} \geq 1$. This enforces that at least one basic area b that is adjacent to a basic area $b' \in S$, but not contained in S, must also assigned to district center i. As a result, these constraints rule out that a district induces a connected component in NG which is not adjacent to the district center i. In case, $\sum_{b \in S} x_{bi} < |S|$ the constraint is trivially fulfilled. By Constraints (5), a driver type is assigned to each delivery district. Constraints (6) ensure that at most the available number of drivers of type $d = 1$ is used. Through Constraints (7) and (8) we make sure that we use at most the number of available drivers of each type $d > 1$, and that driver type priorities are respected. Constraints (9) assign a vehicle type and a number of trips to each delivery district. Constraints (10) guarantee that the available number of vehicles of type $d = 1$ is not exceeded. Analogously to (7) and (8), Constraints (11) and (12) ensure that at most the number of available vehicles of each type $v > 1$ is used and that vehicle type priorities are taken into account. Constraints (13) make sure that vehicle capacities are not exceeded. Constraints (14) limit the number of trips performed to each delivery district to the number of trips required to transport the average daily weight of the district. These constraints are necessary to prevent the model from artificially increasing the workload in a district by making more trips to a district than would be necessary in order to satisfy the workload lower bound. ϵ represents a number slightly greater than zero. Hence, the right-hand sides of these constraints correspond to a ceiling function applied to the average daily weight to be transported to the district, divided by the vehicle capacity. Note that, depending on the choice of ϵ, it may be necessary to round the quotients $\frac{\bar{l}_b}{C_v}$ to a sufficiently small number of decimal places to ensure the correctness of the constraints. Constraints (15) and (16) limit the average daily workload estimation of each district to the interval $[\frac{r_d}{100} LB, \frac{r_d}{100} UB]$ for the driver d assigned to the district. The average daily workload estimation of a district consists of the average workload estimations for the assigned basic areas and the time required to travel between the depot and the district (including reloading the vehicle). The latter results from the minimum number of trips needed to transport the average daily weight of the district. Finally, Constraints (17)–(21) define the binary decision variables.

4.1.2 Model A/IV–AW

Defining the time-expanded binary decision variables z^{τ}_{nvi}:

$$z^{\tau}_{nvi} = \begin{cases} 1 & \text{if vehicle type } v \in V \text{ is assigned to the district represented by basic area } i \in B \\ & \text{and } n \in N \text{ trips to the district are performed on day } \tau \in T \\ 0 & \text{otherwise} \end{cases}$$

we formulate the model as follows:

$$\sum_{b \in B} \sum_{i \in B} c_{bi} x_{bi} \rightarrow \min \tag{22}$$

s.t. (2)–(8), (17)–(20)

$$\sum_{n \in N} \sum_{v \in V} z^{\tau}_{nvi} = x_{ii} \qquad i \in B, \tau \in T \tag{23}$$

$$\sum_{n \in N} z^{\tau}_{nvi} = \sum_{n \in N} z^{1}_{nvi} \qquad v \in V, i \in B, \tau \in T, \tau > 1 \tag{24}$$

$$\sum_{n \in N} \sum_{i \in B} z^{1}_{n1i} \leq N_1 \tag{25}$$

$$\sum_{n \in N} \sum_{i \in B} z^1_{nvi} \leq N_v f_{v-1} \qquad v \in V, v > 1 \tag{26}$$

$$N_v f_v \leq \sum_{n \in N} \sum_{i \in B} z^1_{nvi} \qquad v \in V, v < |V| \tag{27}$$

$$\sum_{b \in B} l^\tau_b x_{bi} \leq \sum_{n \in N} \sum_{v \in V} n C_v z^\tau_{nvi} \qquad i \in B, \tau \in T \tag{28}$$

$$n z^\tau_{nvi} \leq \sum_{b \in B} \frac{l^\tau_b}{C_v} x_{bi} + (1 - \epsilon) \qquad n \in N, v \in V, i \in B, \tau \in T \tag{29}$$

$$LB \sum_{d \in D} \frac{r_d}{100} y_{di} \leq \sum_{b \in B} \overline{w}_b x_{bi} + \frac{1}{|T|} \sum_{n \in N} \sum_{v \in V} \sum_{\tau \in T} t_{ni} z^\tau_{nvi} \qquad i \in B \tag{30}$$

$$UB \sum_{d \in D} \frac{r_d}{100} y_{di} \geq \sum_{b \in B} \overline{w}_b x_{bi} + \frac{1}{|T|} \sum_{n \in N} \sum_{v \in V} \sum_{\tau \in T} t_{ni} z^\tau_{nvi} \qquad i \in B \tag{31}$$

$$z^\tau_{nvi} \in \{0, 1\} \qquad n \in N, v \in V, i \in B, \tau \in T \tag{32}$$

The Objective Function (22) is the same as in model AV–AW. Constraints (23) make sure that a vehicle type and a number of trips is assigned to each delivery district on each day. Constraints (24) guarantee that for each delivery district the same vehicle type is assigned on each day. Constraints (25)−(32) are the time-expanded analogs of Constraints (10)−(16) and (21). In contrast to model AV–AW, the number of trips to each district is determined for each day individually based on the total weight to be transported.

4.1.3 Model IV–A/IW

We formulate model IV–A/IW as follows:

$$\sum_{b \in B} \sum_{i \in B} c_{bi} x_{bi} \to \min \tag{33}$$

s.t. (2)–(8),(17)–(20),(23)–(30),(32)

$$UB \sum_{d \in D} \frac{r_d}{100} y_{di} \geq \sum_{b \in B} w^\tau_b x_{bi} + \sum_{n \in N} \sum_{v \in V} t_{ni} z^\tau_{nvi} \qquad i \in B, \tau \in T \tag{34}$$

The model differs from model A/IV–AW only in one component: Constraints (34) replace Constraints (31). In contrast to model A/IV–AW, the estimated workload of the districts on each day is bounded from above. For each district and day, this estimation contains the day-specific workload estimations for the assigned basic areas and the day-specific travel times between the depot and the district (including reloading times).

4.2 Stage 2: Operational Reassignment

For a given operational day, we adapt the tactical solution computed in the first stage to the concrete demand realization of that day. Let O^{op} be the set of customer orders on the day, and denote by w_b and l_b the corresponding estimated workload and parcel weight, respectively, for basic area $b \in B$. From the tactical solution we derive the following input data for this operational reassignment: By $\Psi \subset B$ we denote the set of open district centers in the tactical solution and by $\Delta_i \subset B$ the set of basic areas in the district represented by $i \in \Psi$. Further, we use $\delta_i \in D$ and $\nu_i \in V$ to denote the driver type and the vehicle type, respectively, that is assigned to the delivery district represented by center $i \in \Psi$. We denote by $\omega \in \mathbb{N}_0$ the maximum number of basic area assignments that are allowed to change compared to the tactical solution. Finally, we formulate the model for the operational reassignment for a given day as follows:

$$\frac{1}{\sum\limits_{i\in\Psi}\sum\limits_{b\in\Delta_i} c_{bi}} \sum_{b\in B}\sum_{i\in\Psi} (c_{bi}x_{bi}) + w^{max} \to \min \tag{35}$$

$$\text{s.t.} \quad w^{max} \geq \frac{100/r_{\delta_i}}{t^{max}} \left(\sum_{b\in B} w_b x_{bi} + \sum_{n\in N} t_{ni} z_{n\nu_i i} \right) \qquad i \in \Psi \tag{36}$$

$$\sum_{i\in\Psi} x_{bi} = 1 \qquad b \in B \tag{37}$$

$$x_{ii} = 1 \qquad i \in \Psi \tag{38}$$

$$\sum_{b\in \bigcup\limits_{b'\in S}(A_{b'}\setminus S)} x_{bi} - \sum_{b\in S} x_{bi} \geq 1 - |S| \qquad i \in \Psi, S \subseteq B \setminus (\{i\} \cup A_i), S \neq \emptyset \tag{39}$$

$$\sum_{n\in N} z_{n\nu_i i} = 1 \qquad i \in \Psi \tag{40}$$

$$\sum_{b\in B} l_b x_{bi} \leq \sum_{n\in N} n C_{\nu_i} z_{n\nu_i i} \qquad i \in \Psi \tag{41}$$

$$n z_{n\nu_i i} \leq \sum_{b\in B} \frac{l_b}{C_{\nu_i}} x_{bi} + (1-\epsilon) \qquad n \in N, i \in \Psi \tag{42}$$

$$|B| - \sum_{i\in\Psi}\sum_{b\in\Delta_i} x_{bi} \leq \omega \tag{43}$$

$$w^{max} \geq 0 \tag{44}$$

$$x_{bi} \in \{0,1\} \qquad b \in B, i \in \Psi \tag{45}$$

$$z_{n\nu_i i} \in \{0,1\} \qquad n \in N, i \in \Psi \tag{46}$$

In Objective Function (35), we aim at optimizing the sum of two terms. The first term reflects geographical compactness and is normalized to a value of approximately one, by dividing by the sum of distances between district centers and assigned basic areas in the tactical solution. The second term represents the maximum workload w^{max} over all districts relative to the contractual working time available in each district, or, for short, the maximum relative workload. The contractual working time that is available in each district is predetermined through the driver type assigned to the district and can be computed as $r_{\delta_i}/100 \cdot t^{max}$. The motivation for the second objective is twofold: First, minimizing the maximum relative workload reduces overtime. Second, it leads to an improvement in the workload balance between the drivers. Note that the maximum relative workload typically takes values close to one. Hence, we treat the two objectives as equally important.

The constraints of the model have the following meaning. Constraints (36), in conjunction with the minimization objective, take care that variable w^{max} is set to the maximum relative workload. Constraints (37) require that each basic area is assigned to a district center. Constraints (38) make sure that the open district centers of the tactical solution remain open. Contiguity is enforced through Constraints (39). Constraints (40) guarantee that just one number n of trips is selected for each district. The vehicle capacity limits are enforced in Constraints (41), and Constraints (42) restrict the number of trips for each district to the number required for the transportation of the district's total weight. Constraints (43) ensure that at most ω basic areas are assigned to a district center different from their center in the tactical solution. Hence, this constraint allows for controlling consistency. Finally, the domain constraints are given by Constraints (44)–(46).

5 Solution Approach

In this section, we describe our approaches to solve the tactical and the operational stage. In preliminary tests, we were able to solve the first stage model for instances with 80 basic areas to reasonable optimality gaps (5.0% on average) within ten hours, using the MIP solver Gurobi 8.1.0. However, for the instances of our case study, which comprise over 250 basic areas, we failed to find even feasible solutions within the time limit of ten hours with models A/IV–AW and IV–A/IW. Only with model AV–AW, we found feasible solutions within the time limit (average optimality gap of 10.8%). Therefore, we opted to devise a heuristic solution approach for all models. The heuristic, which is based on the work of Hess et al. [1965], decomposes the problem into two subproblems, which are then solved iteratively until the solution converges. The first subproblem consists of locating a given number p of district centers. We denote the set of these centers as $I \subseteq B$. The second subproblem then deals with the allocation of basic areas and driver types to centers, where we restrict the set of district centers in each iteration to the set I. Since, in contrast to the work of Hess et al. [1965], we do not know a priori the number of required districts, our models can decide to use only a subset of the centers I and, hence, establish fewer than p districts. Consequently, we refer to the set I as *potential district centers*. The iterative procedure is illustrated in Figure 2. In the following, we address each step individually in detail.

1. Determine the number of potential district centers We determine the number of potential district centers p in a preprocessing step. On the one hand, p should be as small as possible to minimize the computational burden. On the other hand, p must be large enough such that all demand can be accommodated. Since human planners typically know from experience the rough number of districts needed for a particular service region, a good choice is to set p to a value slightly greater than the human planner's estimate. Another way would be to solve vehicle routing problems for a set of sample days, and set p to a value slightly greater than the number of vehicles needed to serve the customers on each sample day, which is similar to the procedure in Schneider et al. [2015].

2. Initialize district centers We use the seeding technique proposed in Arthur and Vassilvitskii [2007] in the context of cluster analysis, to obtain p initial district centers that are evenly spread across the service region. We select the first center uniformly at random. Afterwards, a basic area

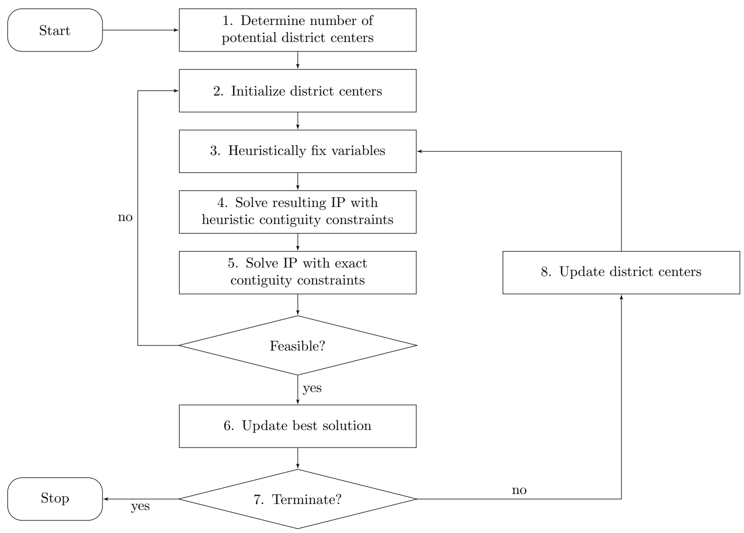


Figure 2: Flowchart of heuristic solution approach

is chosen as an additional center with a probability that is proportional to the squared distance between the basic area and the nearest center already selected. The three models are then adapted accordingly, essentially replacing the expression "$i \in B$" by "$i \in I$".

3. Heuristically fix variables We use the approach presented in Ríos-Mercado and López-Pérez [2013] to heuristically eliminate some of the x_{bi} variables, by forbidding assignments of basic areas to centers that are far away. A basic area b is deemed to be too far away from a center i, if the total workload to serve all basic areas that are closer to i than b exceeds a given threshold $w_{thresh} = \alpha \cdot t^{max}$ with $\alpha \in (0, \infty)$.

4. Solve resulting IP with heuristic contiguity constraints To speed up the algorithm, we first solve the models with the heuristic contiguity constraints proposed in Mehrotra et al. [1998], instead of the exact constraints. These heuristic constraints enforce that for each basic area b assigned to a center i, one of the immediate predecessors of b on a shortest path to i in the neighborhood graph NG is also assigned to i. This condition is more restrictive than the one in Constraints (4), where we just require that some basic area b' adjacent to b is also assigned to i. Thus, we term them as heuristic, as they render some contiguous districts infeasible (although such districts will often not be very compact).

5. Solve IP with exact contiguity constraints In this step, we solve the models with the exact contiguity constraints (4), using the solutions obtained in Step 4 as a warm-start. Due to their exponential number, we add them in a cutting plane fashion, as done in Drexl and Haase [1999]. We can easily separate violated constraints by checking whether each district of an integer solution is contiguous, and if not, adding an exact contiguity constraint for each connected component that does not contain the center (see Ríos-Mercado and López-Pérez [2013]). In addition, as suggested in Salazar-Aguilar et al. [2011], we strengthen our models by adding constraints that prevent single basic areas from being disconnected from their districts. That is, we add all constraints (4) for the special case $|S| = 1$. If no feasible solution can be found in this step, the algorithm continues with Step 2 and probabilistically selects new district centers.

6. Update best solution We have no guarantee that the objective value improves from one iteration of the heuristic to the next, as a solution of a certain iteration is not necessarily feasible in the subsequent iteration, due to the relocation of district centers. Hence, we check if the solution of the current iteration is better than the best solution found so far; and if yes, then we update the best solution with the solution of the current iteration.

7. Terminate The heuristic terminates if one of the following conditions is met: (1) The maximum number of iterations $iter_{max}$ is reached. (2) Cycling is observed, i.e., a solution is obtained in the current iteration that has already been found in a previous iteration.

8. Update district centers For each potential center $i \in I$ that is selected in Step 4 and 5, we select from all basic areas that are assigned to it, the one which, when picked as new center, yields the smallest contribution to the compactness measure of the three models. The new center then replaces the old one in I; centers not being chosen remain unchanged.

If we come to the second stage, we solve the operational reassignment problem based on the tactical solution derived in the first stage and the concrete demand realization of the day. For that, we use the same approach as described in Step 5 of the heuristic. We solve the model for different values of ω to obtain several solutions for an operational day, each with a different emphasis on consistency.

6 Evaluation

We will asses the quality of the solutions computed in the second stage with a given value of ω using four different evaluation measures: the number of districts, driver consistency, workload balance, and operational feasibility. For the calculation of the latter two, we determine the actual workload of each district for the given set O^{op} of orders on that day, instead of relying on the estimates that were used in the two planning stages. While the total service time in each district can simply be calculated as the sum of the service times over all customer orders in that district, we need to solve a routing problem for each district in order to obtain the actual travel time. Although we allow a vehicle to make several trips on a day, we do not consider time restrictions, such as time windows or maximum driving times. Hence, we can compute the actual travel time for each district, including the depot, by solving a capacitated vehicle routing problem (CVRP), see, e.g., Semet et al. [2014]. Vehicle capacities correspond to the loading capacity, and the number of vehicles equals the minimum number of trips required to transport the total weight of the district.

Next, we present the four evaluation measures:

- *Number of districts* (*ND*). This measure describes the number of open district centers, i.e.,

$$ND = |\Psi|.$$

- *Driver consistency* (*DC*). Driver consistency reflects the percentage of customer orders that are carried out by the driver who is intended to serve the corresponding basic area according to the tactical solution. Let $\Delta_i^* \subset B$ denote the set of basic areas that are assigned to the district represented by center $i \in \Psi$ in the operational solution, i.e., *after* the operational reassignment. Then, this measure is computed as

$$DC = \left(1 - \frac{1}{|O^{op}|} \sum_{i\in\Psi} \sum_{b\in\Delta_i, b\notin\Delta_i^*} \Big|\{o \in O^{op} \mid b_o = b\}\Big|\right) \cdot 100[\%].$$

- *Operational feasibility* (*OF*). We define operational feasibility as the percentage of feasible delivery districts—feasible in the sense that the driver in charge does not have to work overtime in order to satisfy the customer demand. Let $t_i^{act} \in \mathbb{R}^+$ denote the actual workload of the district represented by center $i \in \Psi$, i.e., t_i^{act} is equal to the total service and vehicle reloading time plus the actual travel time according to the solution of the corresponding CVRP. Then, this measure is calculated as

$$OF = \frac{1}{|\Psi|} \Big|\{i \in \Psi \mid t_i^{act} \leq \frac{r_{\delta_i}}{100} \cdot t^{max}\}\Big| \cdot 100[\%].$$

- *Workload balance* (*WB*). Workload balance reflects the extent to which the actual workload is balanced evenly between drivers. We denote by $R_i = {(100 \cdot t_i^{act})}/{(r_{\delta_i} \cdot t^{max})}$ the relative workload of the district represented by center $i \in \Psi$. Moreover, we denote by $\mu = \frac{1}{|\Psi|} \cdot \sum_{i\in\Psi} R_i$ the average relative workload over all districts. *WB* is then defined as the maximum absolute deviation between the relative workload of a district and the average relative workload, i.e., it is computed as

$$WB = \max\{\max_{i\in\Psi} R_i - \mu, \mu - \min_{i\in\Psi} R_i\} \cdot 100[\%].$$

  Thus, if $WB = 0$, then the workload is perfectly balanced.

7 Real-World Case Study

In this section, present a case study based on a real-world data set of a European parcel delivery company. In Section 7.1, we briefly describe the underlying data as well as its preparation for the experiments of the case study, and we report the parameterization used in the experiments. We explain how we estimate the travel time within the districts based on the available customer order data (Section 7.2). After that, we experimentally investigate the impact of the following aspects: The values of the workload limits used in the three tactical planning models (Section 7.3), the presence of homogeneous and heterogeneous resources (Sections 7.4.1 and 7.4.2), the location of the depot (Section 7.4.3), and the length of the tactical planning horizon (Section 7.5). We report the running times for both planning stages in Section 7.6 and, finally, visualize some solutions obtained after the operational adaptation in Section 7.7.

7.1 Data Preparation and Parameterization

Our data set comprises approximately 67,000 customer orders delivered in a service region in Germany within a time period of four months. For each customer order $o \in O$, the data includes the day of delivery (τ_o), the service time (s_o), the weight (l_o), and the delivery address. We geocoded the delivery addresses using the PTV xLocate Server[2], and we calculated travel times based on the road network using the PTV xDima Server[2] of our industry partner PTV. The set of basic areas B in the case study corresponds to sub-zip code areas ("PLZ8 areas") provided by PTV. We mapped the customers to their corresponding sub-zip code area, using the free geographic information system QGIS[3]; based on this mapping, we calculated the aggregated service time s_b^τ and weight l_b^τ for each basic area and day. Furthermore, we used QGIS to determine whether two sub-zip code areas share a common border, and then we used this information to derive the neighborhood graph NG and the adjacency information A_b. In Figure 3, we depict the service region and its subdivision into 252 sub-zip code areas. The black triangle represents the depot.

Figure 3: Depot and sub-zip code areas of the service region under study

We split the data set into two separate test instances to account for seasonal demand fluctuations in the data. The first instance comprises the first two months of the data set with roughly 32,000 customer orders, while the second instance contains the remaining two months with approximately 35,000 customer orders. Recall that the data basis for the tactical planning stage consists of a set of sample days representing the expected demand within the planning horizon. To reflect

[2] http://xserver.ptvgroup.com/en-uk/home/ptv-xserver-en/
[3] http://qgis.osgeo.org

this, each instance is, in turn, further subdivided: The first month comprises the set of tactical sample days T that represent the expected future demand and will be the basis for the tactical models, and the second month yields a set of operational days T^{op} that are used in the adjustment stage and the evaluation phase.

For the base case analyzed in Sections 7.1–7.3, we consider the following experimental setup. We assume that a homogeneous fleet is available with capacity $C_v = 1150$ kg, corresponding to a vehicle of the Mercedes Sprinter class, which is the prevalent vehicle class used at the parcel delivery company. Moreover, we consider only full-time drivers, i.e., drivers with $r_d = 100$, and a maximum contractual working time of $t^{max} = 7.5$ hours. The maximum number of trips to a district on the same day is restricted to $|N| = 3$, and reloading a vehicle at the depot takes $t^{reload} = 1/3$ hour. We use $|I| = 16$ potential district centers, which is sufficient to cover the demand of the service region. The number of available drivers M_d and the number of available vehicles N_v are also set to 16.

If we do not state otherwise, we parameterize our solution approach and the evaluation stage as follows. All IP and MIP models presented in this paper are solved using the MIP solver Gurobi 8.1.0 with the following tolerances and time limits:

- For the first stage, we set the MIP optimality tolerance to 3%, which we consider as sufficiently small for practical applications, and the time limit for each of Steps 4 and 5 of our heuristic to 1,800 seconds. We perform at most $iter_{max} = 20$ iterations of the location-allocation procedure and limit the maximum runtime for each instance to 14,400 seconds. Furthermore, we set parameter $\alpha = 3$ for the heuristic variable fixation in Step 3 of the heuristic. In Constraints (14), (29), and (42), we set parameter $\epsilon = 10^{-5}$, which is the solver's default integer feasibility tolerance.

- For the second stage, we set the MIP optimality tolerance to 1% and the time limit for the solution of each MIP model to 60 seconds.

- For the evaluation stage, we set the MIP optimality tolerance to 1%. This means that the evaluation measures that we present in the remainder of this paper were calculated based on near-optimal CVRP solutions.

We implemented all algorithms in Java, and performed all experiments under Ubuntu 16 on a machine with an Intel Xeon E5-2650 v2 CPU at 2.6 GHz and 128 GB of RAM.

7.2 Estimating the Travel Time Within the District

Recall that the estimated workload w_b^τ of basic area $b \in B$ on day $\tau \in T$ within the delivery district consists of the service time s_b^τ plus an estimation for the time required to travel to the customer orders in the basic area on that day. For a given $k \in \mathbb{R}^+$, we compute the travel time estimation for basic area b and day τ as

$$t_b^\tau(k) = \sum_{\substack{o \in O, \\ \tau_o = \tau, b_o = b}} \frac{1}{k} \cdot \left(\sum_{k'=1}^{\lceil k \rceil} t_{o,\kappa(o,k')} - (\lceil k \rceil - k) \cdot t_{o,\kappa(o,k)} \right),$$

where $t_{o_1,o_2} \in \mathbb{R}^+$ denotes the travel time from customer order $o_1 \in O$ to customer order $o_2 \in O$, and $\kappa(o, \epsilon) \in O$ denotes for $\epsilon \in \mathbb{N}^+$ the ϵ-closest customer order from customer order $o \in O$ in terms of travel time (which can be in the given or in another basic area). Thus, for each customer order o, we calculate the average travel time to the k customer orders to which customer order o has the shortest travel time. Note that k does not have to be integer. If k is fractional, we consider the $\lceil k \rceil$-th customer partially. We sum up these values over all customer orders within the basic area on the given day to obtain the travel time estimation $t_b^{\tau}(k)$. This approach is motivated by the observation that it is unlikely that long edges are used in an optimal solution to the vehicle routing problem [see Toth and Vigo, 2003, who develop their granular tabu search based on the same reasoning].

To obtain a value for k which results in a good estimation, we created four test cases: for each of the two test instances of the preceding section, one test case with the original depot and one with a depot centrally located in the service region. We solved the tactical planning problem for each test case with an arbitrary value of $k = 4.5$ using our first-stage solution approach with model A/IV–AW. In the solutions we obtained, we calculated for each $k \in \{1, 1.5, ..., 9.5, 10\}$ the estimated workload for each delivery district (comprising service and reloading time, travel time within the district and between the depot and the district) on each day using the introduced travel time estimation. Additionally, we solved a CVRP for each delivery district on each day to determine the actual workloads. Then, we computed the deviation as well as the absolute deviation between the estimated and the actual workload for each day, district and value of k, and averaged them for each value of k. $k = 4$ yields the best estimation with respect to these two measures, with values of roughly 11 minutes on average for the absolute deviation (see Figure 12 in the appendix for detailed results). Consequently, we use $k = 4$ for all computational experiments in the remainder of this paper.

7.3 Controlling Conservatism

Recall that we expect the three tactical IP models AV–AW, A/IV–AW, and IV–A/IW to be different in terms of how conservative they are. Beyond that, we expect that the degree of conservatism can also be controlled within each model by setting appropriate workload limits LB and UB. In the following, we evaluate both effects. For this purpose, we run experiments on all three models with different workload limits. An overview of all workload limits is given in Table 3. For each model, we consider the three levels (i) LOW, (ii) MEDIUM, and (iii) HIGH, corresponding to upper workload limits UB of 7, 7.5, and 8 hours, respectively. The lower workload limits LB differ, however, between the models. For models AV–AW and A/IV–AW, they correspond to 6, 6.5, and 7 hours, whereas they correspond to only 4.5, 5, and 5.5 hours for model IV–A/IW. As the latter model considers very restrictive workload upper bounds, which prohibit that UB is exceeded on any of the tactical sample days, it is necessary to set relatively low workload limits LB to ensure the feasibility of the model.

The values we obtain with the different workload limits for our evaluation measures on the two test instances are illustrated in Figures 4 and 5. Figures 4a and 5a depict the number of districts. As one would expect, higher workload limits clearly result in the establishment of fewer delivery

Table 3: Overview of different workload limits $[LB; UB]$ for each of the three tactical models (in hours)

	Model		
Workload limit	AV–AW	A/IV–AW	IV–A/IW
LOW	[6; 7]	[6; 7]	[4.5; 7]
MEDIUM	[6.5; 7.5]	[6.5; 7.5]	[5; 7.5]
HIGH	[7; 8]	[7; 8]	[5.5; 8]

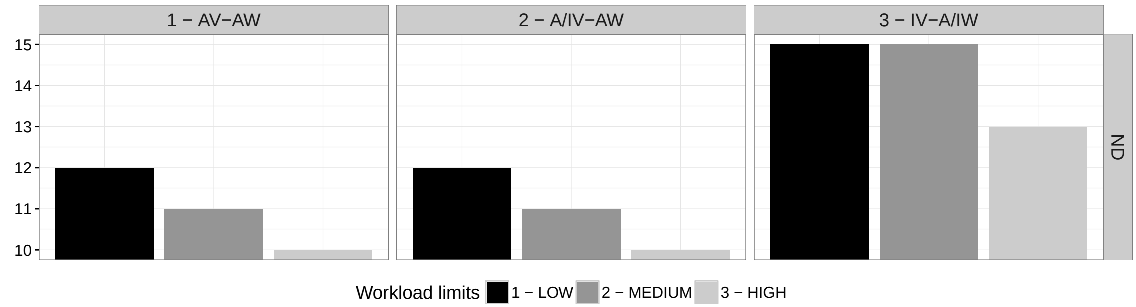
districts. The results also show that model IV–A/IW establishes the highest number of delivery districts of the three models. Moreover, model A/IV–AW tends to establish the same number of delivery districts or one more district compared to model AV–AW.

Figures 4b and 5b show the values for driver consistency, operational feasibility, and workload balance obtained with different values of ω and averaged over all operational days T^{op}. Remember that the parameter ω specifies the maximum number of basic areas that may be reassigned in Stage 2 of the solution approach. Hence, solving each operational problem for each $\omega \in \{0, ..., 20\}$ allows us to evaluate the trade-off between driver consistency and the other evaluation measures.

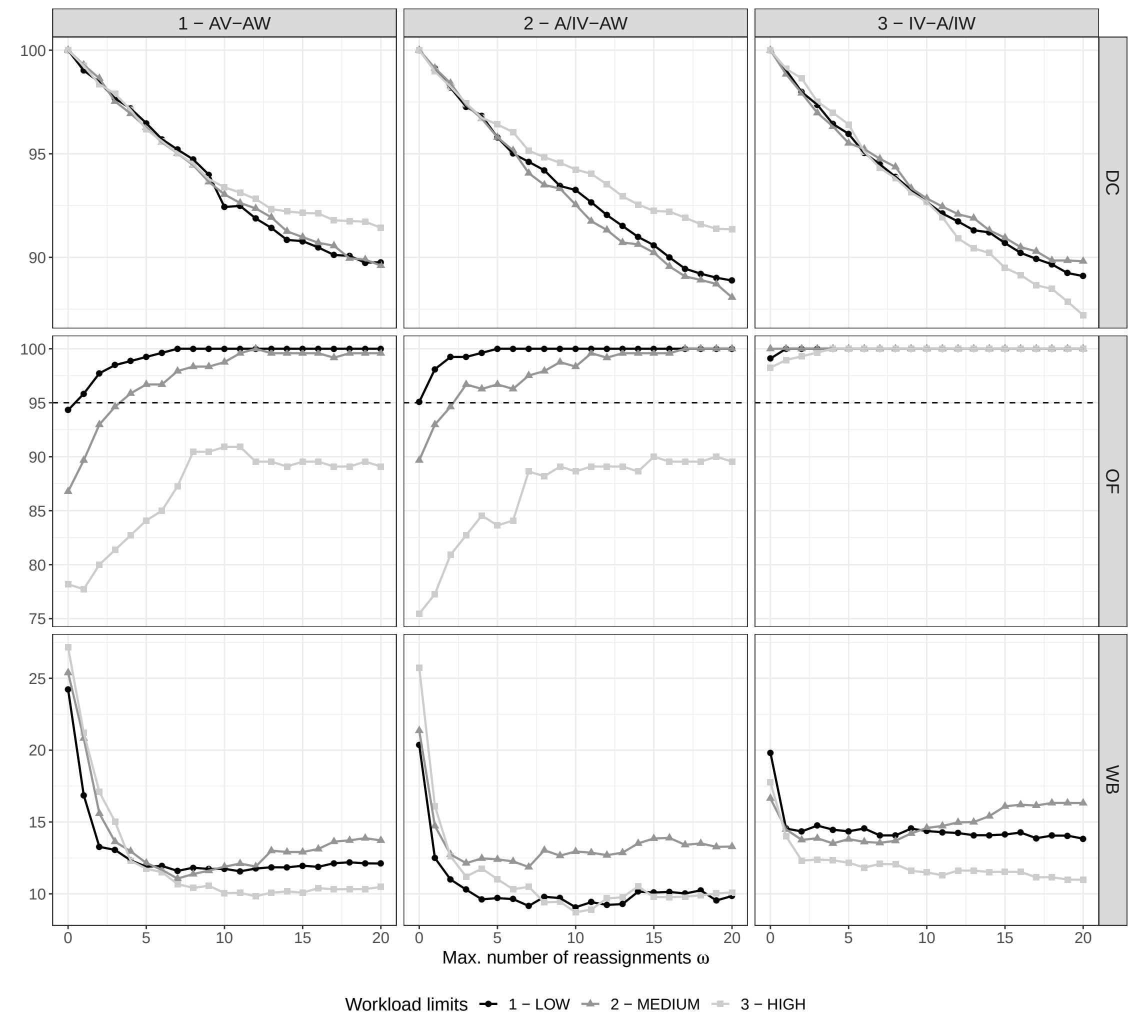
As can be seen from the figures, driver consistency behaves relatively similar for all models and all workload limits, and no consistent differences are discernible. It decreases almost linearly with an increasing value of ω and takes values of approximately 90% for $\omega = 20$.

Major differences, however, can be observed with respect to operational feasibility. Model IV–A/IW clearly provides the best operational feasibility. Irrespective of the level of workload limits, only few reassignments are required to attain values close or equal to 100% on both instances. This confirms the expectation that model IV–A/IW is the most conservative of the three tactical models. The other two models yield significantly lower values for operational feasibility. While the values are very similar for instance 1, for instance 2 there is a discernible difference between the two models. Even for $\omega = 20$, model AV–AW yields an operational feasibility of only approximately 25% and 64% for workload limits HIGH and MEDIUM, respectively. In contrast, model A/IV–AW attains values of about 64% and 94%, respectively. 100% operational feasibility is achieved with the two models only with workload limit LOW. Furthermore, the results show that the degree of conservatism can be controlled by an appropriate choice of workload limits. Suppose, for example, that a human planner targets an operational feasibility of roughly 95%. We marked this value in the figures with a dashed horizontal line. Then, on instance 1, the planner should select the workload limits MEDIUM for models AV–AW and A/IV–AW, and the workload limit HIGH for model IV–A/IW, as these limits result in the fewest number of districts, and, thus, also in the minimum number of required resources, with which the planner's target value is attained. Analogously, the planner should select workload limit LOW for models AV–AW and A/IV–AW, and HIGH for model IV–A/IW on instance 2.

Concerning workload balance, model IV–A/IW yields the best results of the three models for the case that no reassignments are allowed ($\omega = 0$). With an increasing value of ω, one can observe a convergence to fairly similar values for all three models and all workload limits, with model IV–A/IW yielding slightly worse values than the other two models on instance 1.

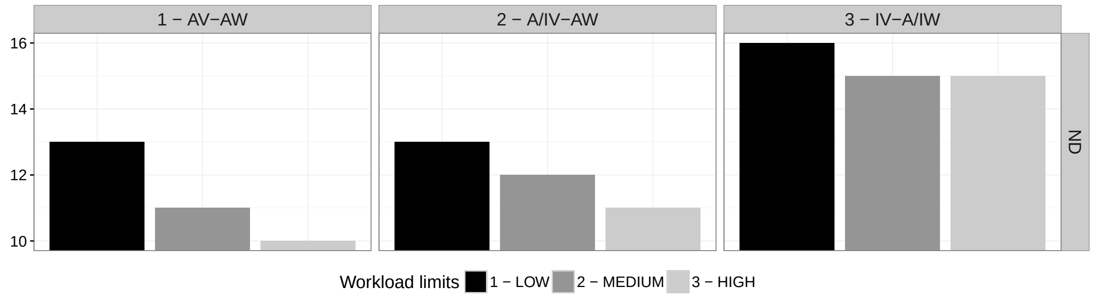


(a) Number of districts

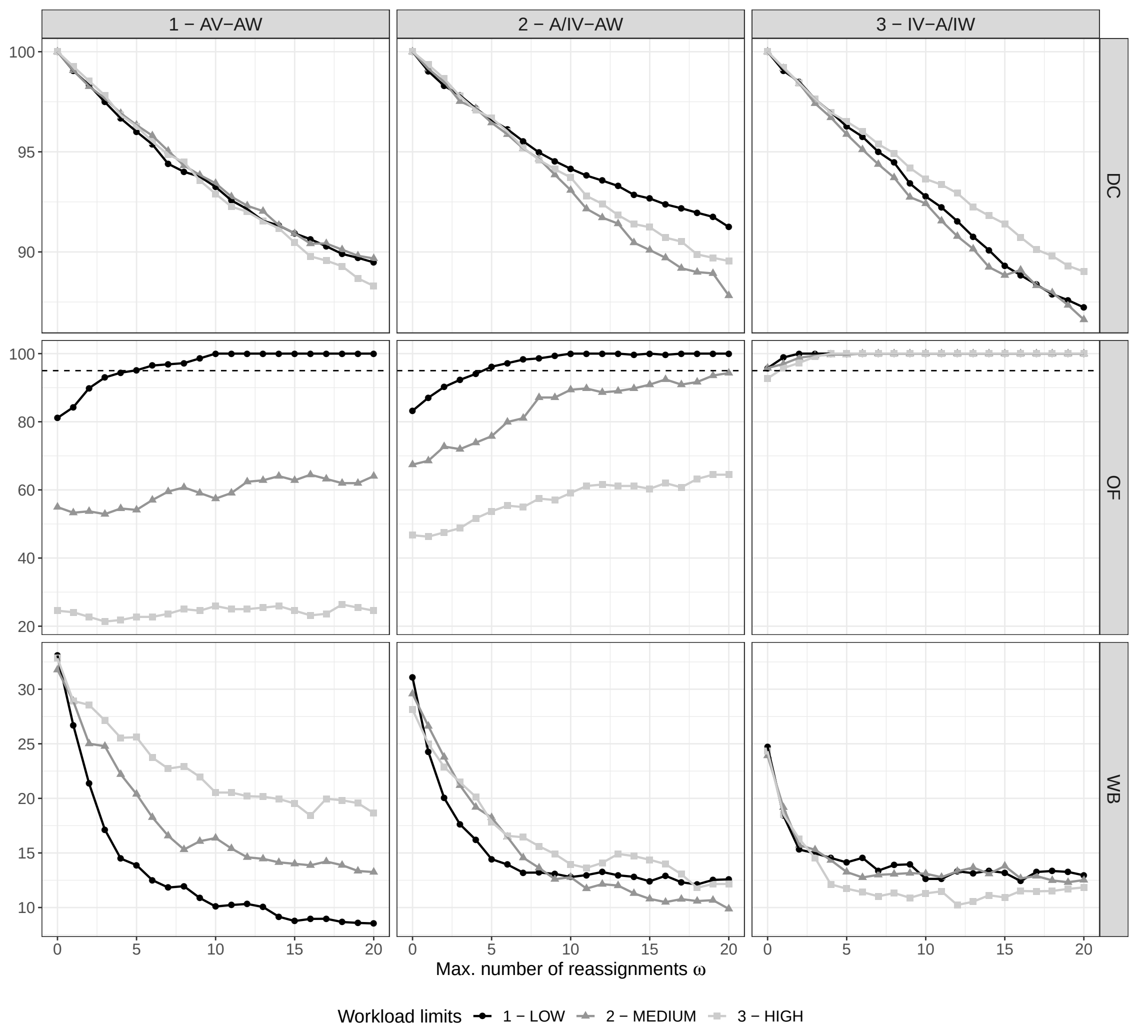


(b) Driver consistency, operational feasibility, and workload balance for different numbers of allowed reassignments (average values over operational days)

Figure 4: Evaluation measures obtained for the three tactical planning models and different workload ranges on test instance 1



(a) Number of districts



(b) Driver consistency, operational feasibility, and workload balance for different numbers of allowed reassignments (average values over operational days)

Figure 5: Evaluation measures obtained for the three tactical planning models and different workload ranges on test instance 2

Apart from the differences between the models and the workload limits, Figures 4b and 5b reveal that instance 1 reaches a significantly better operational feasibility compared to instance 2 if $\omega = 0$. The reason for this lies in differing demand data: In case of instance 1, the average weight per day and the average number of packages per day are slightly higher in the period used for the tactical planning, than in the period used for the operational stage. In instance 2, the situation is the opposite. That means, in this case, the effort per day is, on average, higher than expected, and lower workload limits on the tactical stage and more reassignments on the operational stage become necessary to ensure a high operational feasibility.

All in all, we conclude from these experiments that the models behave in the expected way. The results confirm that the degree of conservatism is influenced by the models themselves and by the choice of the workload limits. It is especially noteworthy that the very high operational feasibility of the three models for the recommended workload limits does not come at the expense of excessively unbalanced work days. On instance 1 (instance 2) the maximal relative deviation from the average workload is below 15% for all $\omega \geq 3$ ($\omega \geq 5$). Hence, with few driver reassignments, our models are able to achieve both balanced workloads across all days and at the same time few or even no days with overtime.

Furthermore, the travel time estimations in the models work quite well, which can be seen from the results for operational feasibility and workload balance: Increasing the value of ω clearly tends to result in an improvement of the two measures, although, in rare cases, it leads to a minor deterioration due to errors in the estimation. In Figure 6, we provide a visual impression of the solutions obtained with model AV–AW and the workload limits recommended above on test instance 1. District boundaries are highlighted by bold lines. Visualizations of the solutions obtained with models A/IV–AW and IV–A/IW can be found in Figures 14 and 15 of the appendix.

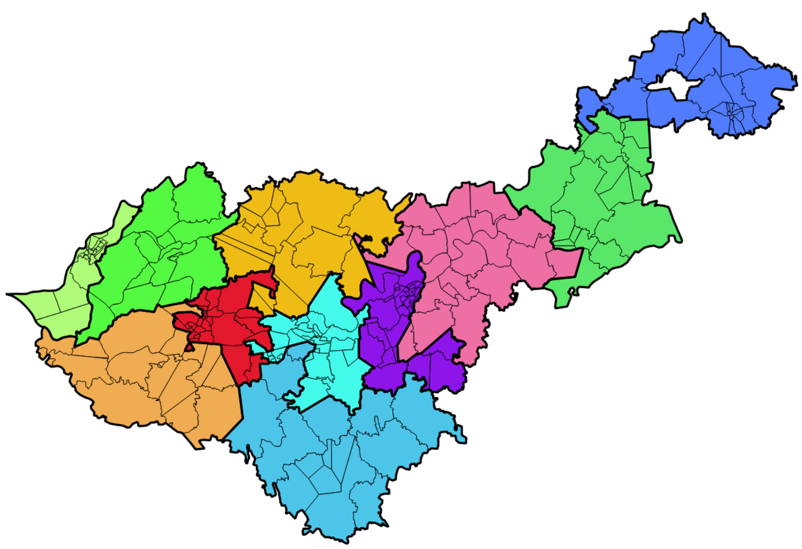


Figure 6: Tactical district design obtained with model AV–AW on instance 1

If a planner targets 95% operational feasibility, we recommend configuring the tactical models as follows: For models AV–AW and A/IV–AW, workload limits MEDIUM on instance 1 and LOW on instance 2 should be used. For model IV–A/IW, the workload limit HIGH should be used on both instances. If we take the minimum value of ω required to reach 95% operational feasibility on average, we also observe satisfying results on an individual level over all models: In the worst case, the contractual working time of a driver is exceeded only in four out of 22 working days. And the maximum overtime for a driver was about 2 hours.

For the remainder of this paper, we fix the workload limits of the three models according to our recommendations.

7.4 Resources and Depot Configuration

In the following, we examine the effect of different resources and depot configurations. We start by introducing a heterogeneous crew of drivers and a heterogeneous fleet of vehicles. Subsequently, we analyze the impact of having a depot that is centrally located in the service region. Each effect is studied individually, i.e., in each of the following subsections the parameterization changes only in one aspect compared to the parameterization described in Sections 7.1–7.3.

7.4.1 Different Driver Types.

So far, we assumed that the crew of available drivers consists only of full-time drivers. In this section, we extend the crew of available drivers to two different driver types: We consider ten full-time drivers ($M_1 = 10$, $r_1 = 100$) and six drivers with a contractual working time of 75% ($M_2 = 6$, $r_2 = 75$), with full-time drivers being prioritized.

In Figure 7, we illustrate the number of districts that we obtain with this crew of drivers (heterogeneous drivers) and compare the results with the number of districts established for the case that only full-time drivers are available (homogeneous drivers). We report these numbers per model and per test instance.

(a) Instance 1

(b) Instance 2

Figure 7: Number of districts obtained with the three tactical planning models for a heterogeneous and a homogeneous crew of available drivers

The figure shows that the number of districts increases by one or two for all models when a heterogeneous crew of drivers is considered. This is due to the fact that some districts must be served by part-time drivers, who can only cover a smaller area. As a result, the total number of drivers and districts increases.

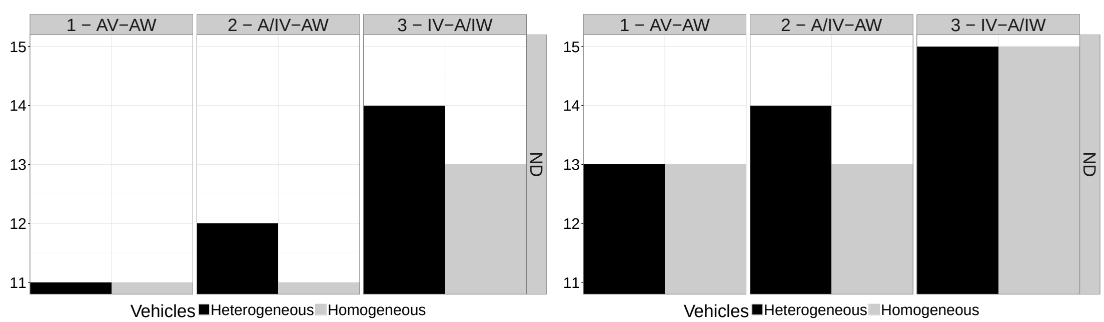
Analogous to Section 7.3, we computed for each model, test instance and value of ω the absolute deviation in the average daily values obtained for measures driver consistency, operational feasibility, and workload balance between the case of a homogeneous crew and the case of a heterogeneous crew. The 90%-quantile for the absolute deviations amounts to 1.9%, 5.8%, and 4.2% for driver consistency, operational feasibility, and workload balance, respectively. Since the values deviate only by a few percentage points from those reported in Section 7.3 for a homogeneous crew of drivers, we omit the figures for these measures.

The solutions obtained with a heterogeneous crew on test instance 1 are depicted in Figures 16–18 in the appendix.

7.4.2 Different Vehicle Types.

Now we consider the case that a heterogeneous fleet of vehicles is available. More precisely, we assume that we have $N_1 = 10$ standard vehicles with capacity $C_1 = 1150$ kg and $N_2 = 8$ small vehicles with capacity $C_2 = 800$ kg. Accordingly, the number of potential district centers is set to $|I| = 18$.

Figure 8 shows the resulting number of delivery districts in comparison to the case of a homogeneous fleet. Although the relative difference in vehicle capacity between a small and a standard vehicle is greater than the relative difference in working time between a part-time and a full-time driver as considered in the preceding section, the increase in the number of districts is smaller. For half of the cases, we obtain the same number of districts, and only one additional district is established in all other cases. Hence, the vehicle capacity seems to be a less restrictive factor than working time on these instances.



(a) Instance 1

(b) Instance 2

Figure 8: Number of districts obtained with the three tactical planning models for a heterogeneous and a homogeneous fleet of vehicles

We refrain from reporting the values for measures driver consistency, operational feasibility, and workload balance due to their similarity with those reported in Section 7.3 for a homogeneous fleet: The 90%-quantile for the absolute deviations obtained for these measures amounts to 2.0%, 6.2%, and 2.7%, respectively.

The solutions obtained with a heterogeneous fleet on instance 1 are illustrated in Figures 19–21 in the appendix.

7.4.3 Location of the Depot.

In the following, we investigate the impact of the depot location. Recall that the original depot is fairly remote from the service region (see Figure 3). We compare this now with a setting where the depot is centrally located in the service region (see Figure 13 in the appendix for the exact location of the depot). Again, all other parameters are set to the values described in Section 7.3.

Figure 9 contains the results obtained for the two depot configurations. The values for driver consistency, operational feasibility, and workload balance deviate only slightly from the numbers of Section 7.3: The 90%-quantiles of the absolute deviations with respect to the values obtained for the original depot location are 2.1%, 3.1%, and 4.3%, respectively. Hence, we exclude these measures from the figure.

The Figures 9a and 9b show the number of districts generated by the three models. If the depot is located centrally, the number of districts can be reduced by at least two with all models, due to the shorter travel time between the depot and the delivery districts.

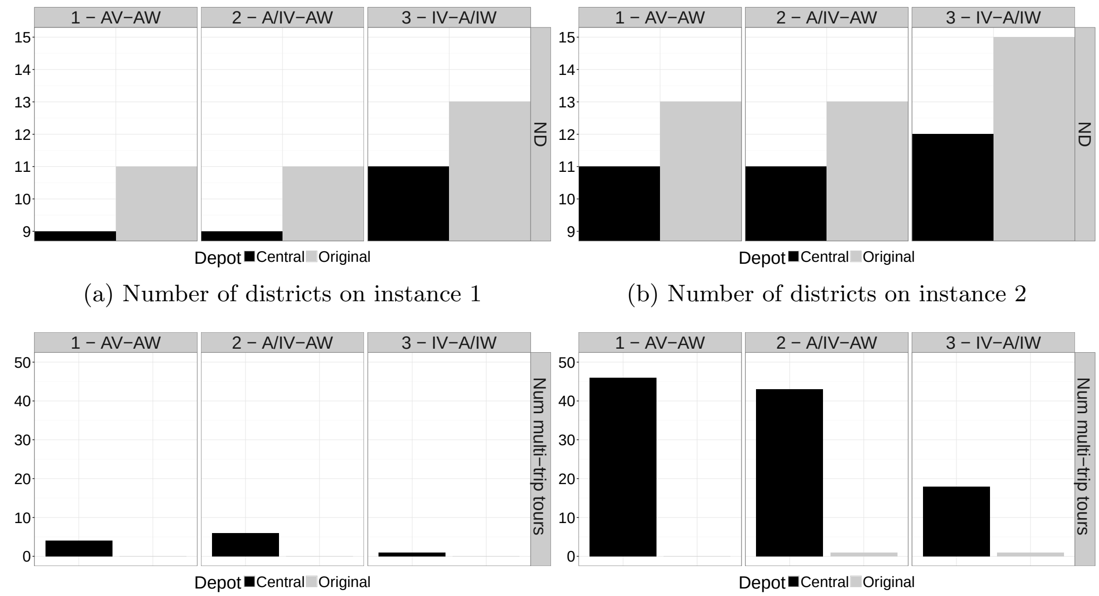
Figures 9c and 9d show the total number of multi-trip tours performed on the operational days for $\omega = 20$. With the original depot, the models try to completely avoid multi-trip tours, since the depot's remote location leads to a large increase in workload for each additional trip to a delivery district. However, with a central depot, multi-trip tours become more attractive, as the additional travel time between the depot and the delivery districts is drastically reduced and, thus, the increase in workload is only moderate, in particular for those districts directly surrounding the depot. The fact that more multi-trip tours are performed for test instance 2 than for test instance 1 can be explained by a considerably higher total weight that must be transported in test instance 2.

7.5 Length of the Tactical Planning Horizon $|T|$

Next, we perform a sensitivity analysis with respect to the length of the planning horizon $|T|$ considered in the three tactical planning models. We compare the results we obtain with planning horizons consisting of one week, two weeks, and an entire month. Figure 10 contains the number of districts and the values for operational feasibility. We omit again the values for driver consistency and workload balance. The 90%-quantile for their absolute deviations with respect to the numbers reported in Section 7.3 equals 2.2% and 8.1%, respectively.

On instance 1, the selection of the planning horizon does hardly influence the results: The number of resulting districts stays the same for all models, and the operational feasibility barely shows differing results for different values of $|T|$.

However, in case of instance 2, the planning horizons of one and two weeks seem to be too short, especially in case of the less conservative models AV–AW and A/IV–AW. For these cases, an operational feasibility of 95% cannot be reached even with high values of ω. Considering the data of one month, one more district is created, and operational feasibility can be achieved with few reassignments.



(a) Number of districts on instance 1

(b) Number of districts on instance 2

(c) Number of multi-trip tours on instance 1 for $\omega = 20$ (sum over operational days)

(d) Number of multi-trip tours on instance 2 for $\omega = 20$ (sum over operational days)

Figure 9: Number of districts, operational feasibility and number of multi-trip tours obtained for the three tactical planning models and different depot locations

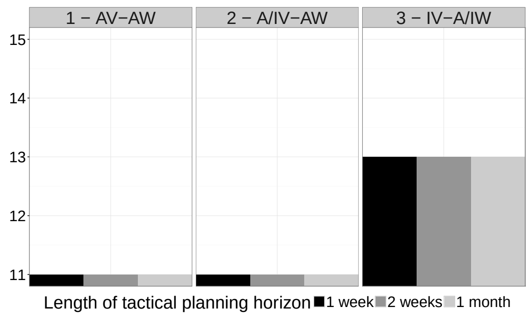
If a planning horizon of one month is selected, the most conservative model IV–A/IW even establishes two more districts on instance 2. This can be explained by the way in which the model handles workload limits: The estimated workload for every single day—even "peak" days—must not exceed a given threshold, and, thus, the model tends to create more districts with an increasing length of the planning horizon. However, we can also achieve an operational feasibility of at least 95% by considering just one or two weeks, and allowing a relatively small number of reassignments.

A general statement on the optimal length of the planning horizon cannot be made. However, we can state that the first two models considering the workloads as average, are less sensitive to demand fluctuations and trends, and, in tendency, should be executed on longer planning horizons. In presence of high demand fluctuations, the conservative model IV–A/IW is more likely to provide satisfactory results in terms of operational feasibility, if the planning horizon, for which representative demand days are area available, is rather short.

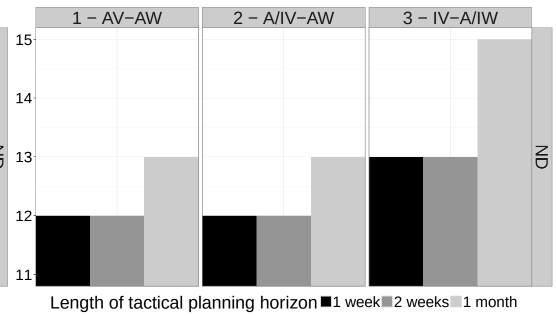
7.6 Running Times

Table 4 contains the running times of the location-allocation heuristic in seconds grouped by the three tactical planning models. We include all experiments presented in the preceding sections with a planning horizon of one month, and report the mean, the minimum, and the maximum running time for each model. Furthermore, we report the mean number of location-allocation iterations performed.

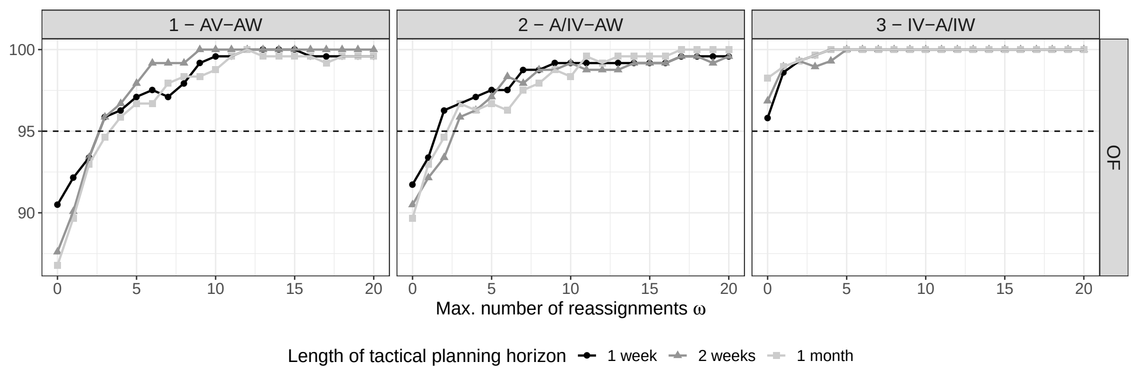
Since model AV–AW does not consider day-specific input data, it contains the smallest number of variables of the three models, which results in the shortest running times. Model A/IV–AW is



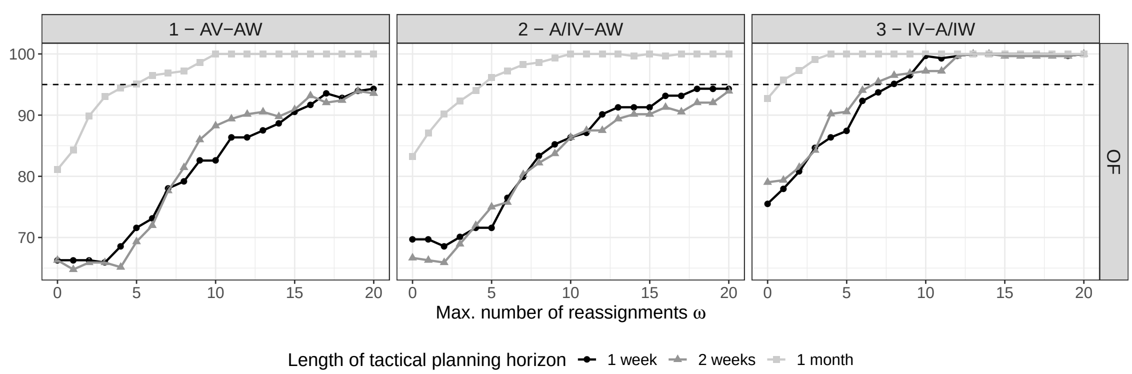
(a) Number of districts on instance 1



(b) Number of districts on instance 2



(c) Operational feasibility for different numbers of allowed reassignments on instance 1 (average values over operational days)



(d) Operational feasibility for different numbers of allowed reassignments on instance 2 (average values over operational days)

Figure 10: Number of districts and operational feasibility obtained for the three tactical planning models and different lengths of the tactical planning horizon $|T|$

the computationally most challenging model. It has the highest average running time, even though the fewest iterations are performed when this model is used.

The location-allocation heuristic addresses a tactical planning problem, which is typically solved only every few months. Hence, the reported running times of a few seconds up to four hours do not pose a limitation on the suitability of the heuristic for practice, irrespective of the underlying model. However, taking into account the uncertainty of demand forecasts, the relatively short running times of the model variant AV–AW can be exploited by the planner to evaluate different demand scenarios for performing what-if analyses.

For the operational reassignment model, we obtain an average running time of 7.4 seconds and an average optimality gap of 1.04%. Hence, this model is well suited for day-to-day planning and even can be used by planners to "play" with different values of ω.

7.7 Visualization of Operational Reassignments

Figure 11 exemplarily illustrates the operational reassignments made for $\omega = 10$ on a particular operational sample day. The underlying tactical district design was computed with model A/IV–AW on test instance 1. Tactical district boundaries are marked by bold lines, the districts resulting from the operational adaptation are distinguished by different colors. Reassigned basic areas compared to the tactical solution are highlighted by diagonal lines. Figures 22–24 in the appendix show the reassignments for $\omega \in \{5, 15, 20\}$.

8 Conclusions

In this paper, we have studied a real-world problem arising in parcel delivery and presented a solution framework that relies on districting approaches that yield well-balanced, compact, and operationally feasible tour plans on a day-to-day basis with a high degree of driver consistency. To the best of our knowledge, we are the first to address a districting problem that integrates the determination of the number of districts and the assignment of heterogeneous resources to districts. This enables the planner to easily assess the impact of different combinations of vehicle and driver types.

Corresponding to the two-stage nature of the problem, we have presented a two-stage solution approach capable of designing districts on a tactical level and adjusting them in day-to-day operations. Its effectiveness has been shown in an extensive case study on real-world data. The case

Table 4: Running times and number of iterations of the location-allocation heuristic for the three tactical planning models

	Running time [s]			
Model	Mean	Min	Max	Iterations
AV–AW	295	15	1,203	8.2
A/IV–AW	4,694	483	14,400	6.6
IV–A/IW	3,343	435	14,400	9.0

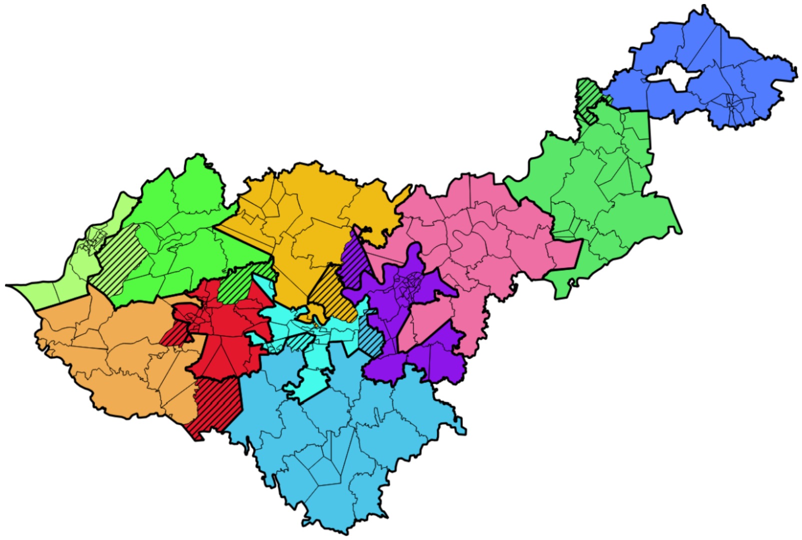


Figure 11: Solution obtained using model A/IV–AW on instance 1 after operational reassignment with $\omega = 10$

study revealed that only few adaptations of the tactical district design are necessary to achieve a high degree of operational feasibility along with a very good workload balance.

Moreover, the case study showed that the three tactical planning models behave as expected. Hence, conservative planners should choose model IV–A/IW since this model produces the best operational feasibility with very good results even if no or only few operational reassignments are allowed. However, the high degree of operational feasibility is achieved at the expense of the highest number of districts of the three models. Less conservative planners and planners willing to accept a slightly higher number of operational reassignments should select model AV–AW or model A/IV–AW, both yielding fairly similar results in the relevant evaluation measures. If computation time is an issue, preference should be given to model AV–AW.

Acknowledgments This work was partly supported by PTV Group. The authors thank PTV for the excellent cooperation and, in particular, for providing map data of sub-zip code areas as well as their PTV xLocate and PTV xDima servers

References

D. Arthur and S. Vassilvitskii. k-means++: The advantages of careful seeding. In *Proceedings of 18th ACM-SIAM Symposium on Discrete Algorithms*, pages 1027–1035. SIAM, 2007.

J. F. Bard and A. I. Jarrah. Large-scale constrained clustering for rationalizing pickup and delivery operations. *Transportation Research Part B: Methodological*, 43(5):542–561, 2009. ISSN 01912615.

J. G. Carlsson. Dividing a territory among several vehicles. *INFORMS Journal on Computing*, 24 (4):565–577, 2012.

J. G. Carlsson and E. Delage. Robust partitioning for stochastic multivehicle routing. *Operational Research*, 61(3):727–744, 2013.

C. F. Daganzo and A. L. Erera. On planning and design of logistics systems for uncertain environments. In M. G. Speranza and P. Stähly, editors, *New Trends in Distribution Logistics. Lecture Notes in Economics and Mathematical Systems*, volume 480, pages 3–21. Springer Berlin / Heidelberg, 1999.

A. Drexl and K. Haase. Fast approximation methods for sales force deployment. *Management Science*, 45(10):1307–1323, 1999.

A. L. Erera. *Design of Large-Scale Logistics Systems for Uncertain Environments*. PhD thesis, University of California, Berkeley, 2000.

B. Fleischmann and J. N. Paraschis. Solving a large scale districting problem: A case report. *Computers & Operations Research*, 15(6):521–533, 1988.

L. C. Galvão, A. G. N. Novaes, J. E. Souza de Cursi, and J. C. Souza. A multiplicatively-weighted Voronoi diagram approach to logistics districting. *Computers & Operations Research*, 33(1): 93–114, 2006. ISSN 03050548.

R. G. González-Ramírez, N. R. Smith, R. G. Askin, P. A. Miranda, and J. M. Sánchez. A hybrid metaheuristic approach to optimize the districting design of a parcel company. *Journal of Applied Research and Technology*, 9(1):19–35, 2011.

C. Groër, B. Golden, and E. Wasil. The consistent vehicle routing problem. *Manufacturing & Service Operations Management*, 11(4):630–643, 2009. ISSN 1523-4614.

M. A. Haughton. The efficacy of exclusive territory assignments to delivery vehicle drivers. *European Journal of Operational Research*, 184(1):24–38, 2008. ISSN 03772217.

D. Haugland, S. C. Ho, and G. Laporte. Designing delivery districts for the vehicle routing problem with stochastic demands. *European Journal of Operational Research*, 180(3):997–1010, 2007. ISSN 03772217.

S. W. Hess, J. B. Weaver, H. J. Siegfeldt, J. N. Whelan, and P. A. Zitlau. Nonpartisan political redistricting by computer. *Operations Research*, 13(6):998–1006, 1965.

C. Holland, J. Levis, R. Nuggehalli, B. Santilli, and J. Winters. UPS optimizes delivery routes. *Interfaces*, 47(1):8–23, 2017.

J. Janssens, J. Van Den Bergh, K. Sörensen, and D. Cattrysse. Multi-objective microzone-based vehicle routing for courier companies: From tactical to operational planning. *European Journal of Operational Research*, 242(1):222–231, 2015. ISSN 03772217.

A. I. Jarrah and J. F. Bard. Large-scale pickup and delivery work area design. *Computers & Operations Research*, 39(12):3102–3118, 2012. ISSN 03050548.

J. Kalcsics. Districting problems. In G. Laporte, S. Nickel, and F. Saldanha da Gama, editors, *Location Science*, pages 595–622. Springer International Publishing, 1 edition, 2015. ISBN 9783319131108.

A. A. Kovacs, S. N. Parragh, and R. F. Hartl. A template-based adaptive large neighborhood search for the consistent vehicle routing problem. *Networks*, 63(1):60–81, 2014.

H. Lei, G. Laporte, and B. Guo. Districting for routing with stochastic customers. *EURO Journal on Transportation and Logistics*, 1(1):67–85, 2012. ISSN 2192-4376.

H. Lei, R. Wang, and G. Laporte. Solving a multi-objective dynamic stochastic districting and routing problem with a co-evolutionary algorithm. *Computers & Operations Research*, 67:12–24, 2016. ISSN 03050548.

P. Matl, R. F. Hartl, and T. Vidal. Workload equity in vehicle routing problems: A survey and analysis. *Transportation Science*, 52(2):239–260, 2017.

A. Mehrotra, E. L. Johnson, and G. L. Nemhauser. An optimization based heuristic for political districting. *Management Science*, 44(8):1100–1114, 1998. ISSN 0025-1909.

A. G. N. Novaes and O. D. Graciolli. Designing multi-vehicle delivery tours in a grid-cell format. *European Journal of Operational Research*, 119(3):613–634, 1999. ISSN 03772217.

A. G. N. Novaes, J. E. S. de Cursi, and O. D. Graciolli. A continuous approach to the design of physical distribution systems. *Computers & Operations Research*, 27(9):877–893, 2000. ISSN 03050548.

Y. Ouyang. Design of vehicle routing zones for large-scale distribution systems. *Transportation Research Part B: Methodological*, 41(10):1079–1093, 2007. ISSN 01912615.

R. Z. Ríos-Mercado and J. F. López-Pérez. Commercial territory design planning with realignment and disjoint assignment requirements. *Omega*, 41(3):525–535, 2013. ISSN 03050483.

D. G. Rossit, D. Vigo, F. Tohmé, and M. Frutos. Visual attractiveness in routing problems: A review. *Computers & Operations Research*, 2018.

M. A. Salazar-Aguilar, R. Z. Ríos-Mercado, and M. Cabrera-Ríos. New models for commercial territory design. *Networks and Spatial Economics*, 11(3):487–507, 2011.

M. Schneider, A. Stenger, F. Schwahn, and D. Vigo. Territory-based vehicle routing in the presence of time-window constraints. *Transportation Science*, 49(4):732–751, 2015. ISSN 0041-1655.

F. Semet, P. Toth, and D. Vigo. Chapter 2: Classical exact algorithms for the capacitated vehicle routing problem. In *Vehicle Routing: Problems, Methods, and Applications, Second Edition*, pages 37–57. SIAM, 2014.

K. Smilowitz, M. Nowak, and T. Jiang. Workforce management in periodic delivery operations. *Transportation Science*, 47(2):214–230, 2013. ISSN 0041-1655.

P. Toth and D. Vigo. The granular tabu search and its application to the vehicle-routing problem. *INFORMS Journal on Computing*, 15(4):333–346, 2003.

T. Vidal, G. Laporte, and P. Matl. A concise guide to existing and emerging vehicle routing problem variants. *European Journal of Operational Research*, 2019.

K. Wong and J. Beasley. Vehicle routing using fixed delivery areas. *Omega*, 12(6):591–600, 1984. ISSN 03050483.

R. T. Wong. Vehicle routing for small package delivery and pickup services. In B. Golden, S. Raghavan, and E. Wasil, editors, *The Vehicle Routing Problem: Latest Advances and New Challenges*, pages 475–485. Springer US, Boston, 2008. ISBN 978-0-387-77778-8.

H. Zhong, R. W. Hall, and M. Dessouky. Territory planning and vehicle dispatching with driver learning. *Transportation Science*, 41(1):74–89, 2007. ISSN 0041-1655.

A Summary of Notation

Table 5 contains the entire notation that is used in the models for tactical districting and operational reassignment. The notation below the dotted lines is used only in the model for operational reassignment.

Table 5: Summary of the notation used in the models of the two-stage solution approach

Sets	
B	Basic areas
$I \subset B$	Potential district centers
T	Tactical sample days
T^{op}	Operational days
O	Tactical customer orders
O^{op}	Operational customer orders
D	Driver types
V	Vehicle types
N	Number of trips per tour
$A_b \subseteq B$	Adjacent basic areas to basic area $b \in B$
$\Psi \subset B$	Open district centers in tactical solution
$\Delta_i \subset B$	Basic areas in district of tactical solution represented by center $i \in \Psi$
Parameters	
$c_{bi} \in \mathbb{R}^+$	Distance between basic areas b and i, b, $i \in B$
$t^{max} \in \mathbb{R}^+$	Contractual working time per day of a full-time driver
$r_d \in (0, 100]$	Relative working time of driver type $d \in D$ in percent
$M_d \in \mathbb{N}^+$	Number of available drivers of type $d \in D$
$C_v \in \mathbb{R}^+$	Capacity of vehicle type $v \in V$
$N_v \in \mathbb{N}^+$	Number of available vehicles of type $v \in V$
$l_b^\tau \in \mathbb{R}^+$	Total weight in basic area $b \in B$ on day $\tau \in T$
$\bar{l}_b \in \mathbb{R}^+$	Average total weight in basic area $b \in B$ per day on tactical sample days
$w_b^\tau \in \mathbb{R}^+$	Estimated total workload (service + travel time) of basic area $b \in B$ on day $\tau \in T$
$\overline{w}_b \in \mathbb{R}^+$	Average estimated total workload of basic area $b \in B$ per day on tactical sample days
$t_{ni} \in \mathbb{R}^+$	Travel time between depot and district represented by basic area $i \in B$ plus reloading time at the depot if $n \in N$ trips to the district are made
LB, $UB \in \mathbb{R}^+$	Lower and upper workload limits
$\delta_i \in D$	Driver type assigned to district represented by center $i \in \Psi$ in tactical solution
$\nu_i \in V$	Vehicle type assigned to district represented by center $i \in \Psi$ in tactical solution
$\omega \in \mathbb{N}_0$	Maximum number of basic areas that may be reassigned compared to tactical solution
Variables	
$x_{bi} \in \{0, 1\}$	Takes a value of 1 if and only if basic area $b \in B$ is assigned to the district represented by center $i \in B$
$y_{di} \in \{0, 1\}$	Takes a value of 1 if and only if driver type $d \in D$ is assigned to the district represented by center $i \in B$
$z_{nvi}^{(\tau)} \in \{0, 1\}$	Takes a value of 1 if and only if vehicle type $v \in V$ performs $n \in N$ trips to the district represented by center $i \in B$ (on day $\tau \in T$)
$e_d \in \{0, 1\}$	Takes a value of 1 if and only if all available drivers of type $d \in D$ are assigned a districts
$f_v \in \{0, 1\}$	Takes a value of 1 if and only if all available vehicles of type $v \in V$ are assigned to districts
$w^{max} \in \mathbb{R}^+$	Maximum relative workload of all districts

B Quality of travel time estimations for different values of parameter k

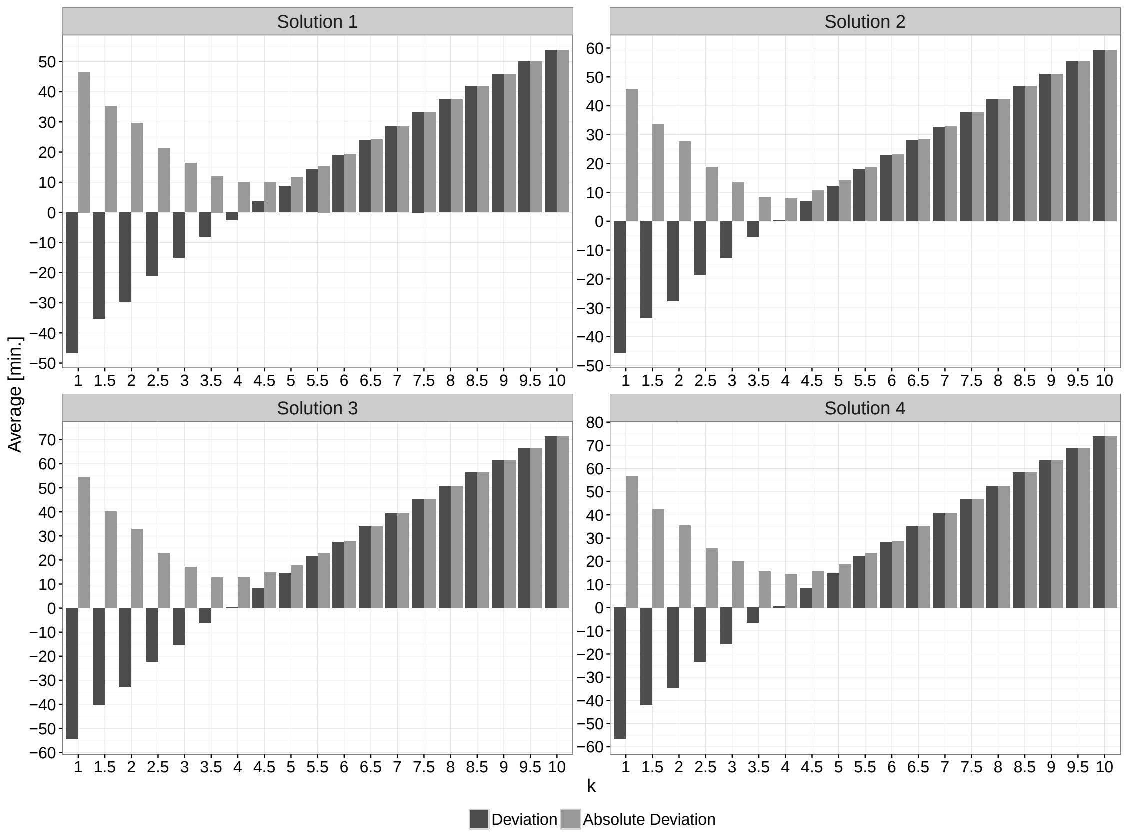


Figure 12: Quality of travel time estimations for different values of parameter k measured as deviation and absolute deviation between estimated and actual workload (average values over districts and days)

C Position of Centrally Located Depot

Figure 13 shows the position of the centrally located depot used for the experiments in Section 7.4.3.

Figure 13: Service region with centrally located depot (represented by the black triangle)

D Visualizations of Tactical District Designs

D.1 Controlling Conservatism

As a supplement to Section 7.3, Figures 14 and 15 show the tactical district designs obtained with the models A/IV–AW and IV–A/IW on test instance 1.

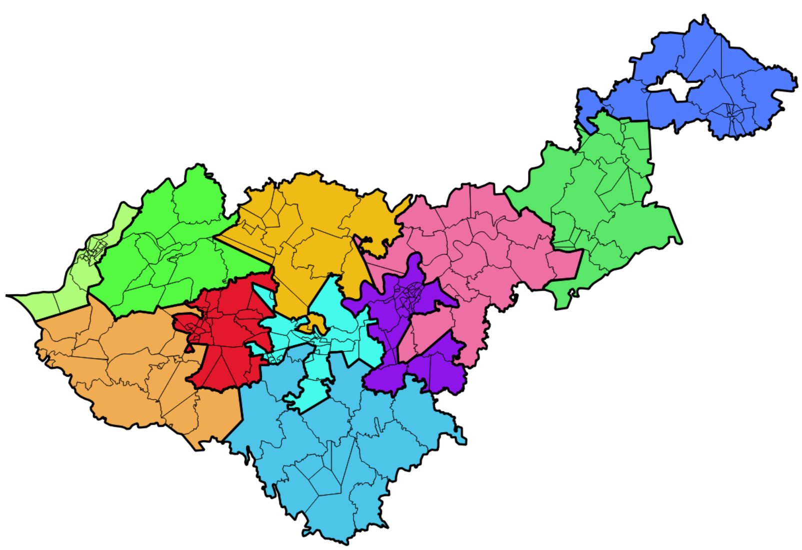


Figure 14: Tactical district design obtained with model A/IV–AW on instance 1

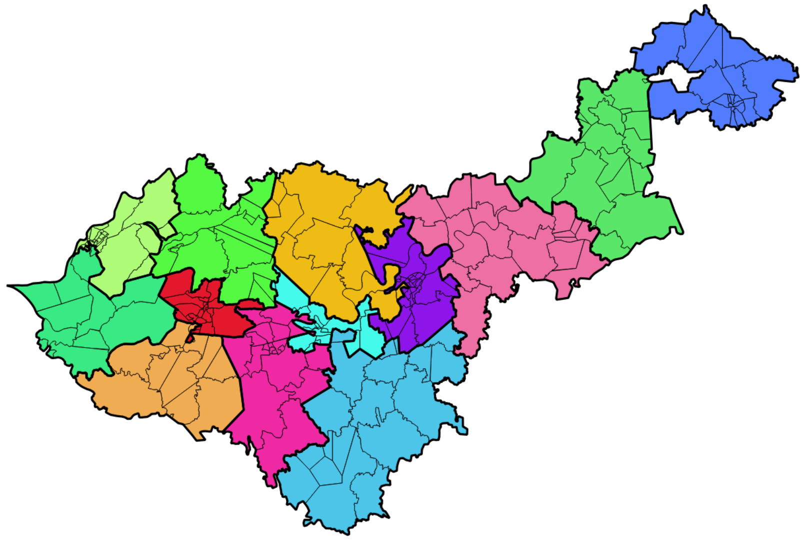


Figure 15: Tactical district design obtained with model IV–A/IW on instance 1

D.2 Different Driver Types

Complementing the results of Section 7.4.1, Figures 16–18 show the tactical district designs obtained with the three tactical planning models on test instance 1 for a heterogeneous crew of drivers. Shaded delivery districts indicate the assignment of a part-time driver.

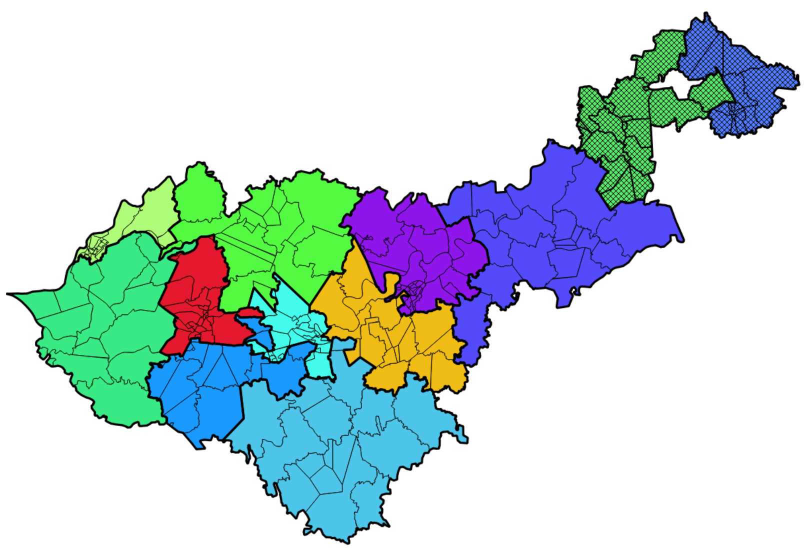


Figure 16: Tactical district design obtained with model AV–AW on instance 1 with a heterogeneous crew of drivers

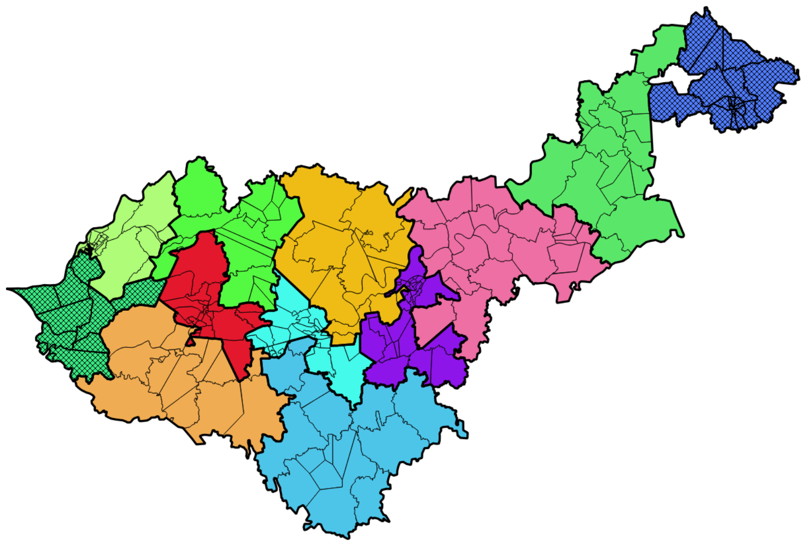


Figure 17: Tactical district design obtained with model A/IV–AW on instance 1 with a heterogeneous crew of drivers

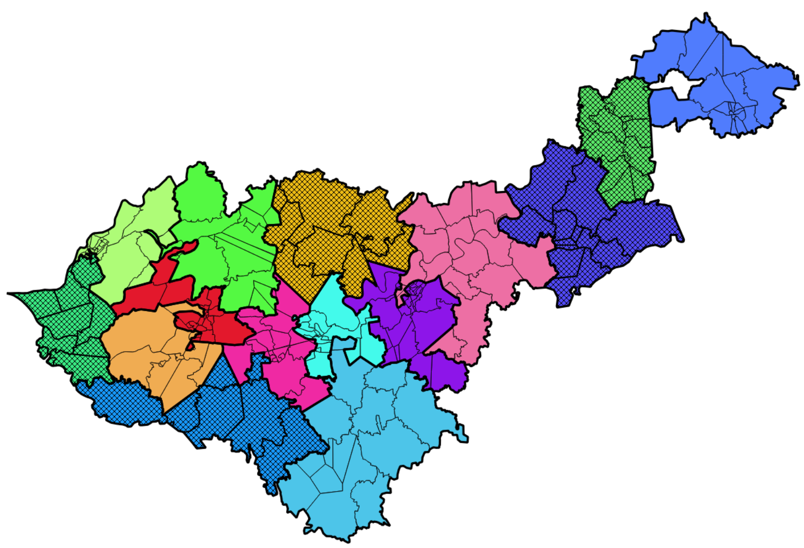


Figure 18: Tactical district design obtained with model IV–A/IW on instance 1 with a heterogeneous crew of drivers

D.3 Different Vehicle Types

Figures 19–21 show the tactical district designs obtained with the three tactical planning models on test instance 1 when a heterogeneous fleet of vehicles is available. Delivery districts with a small vehicle are shaded.

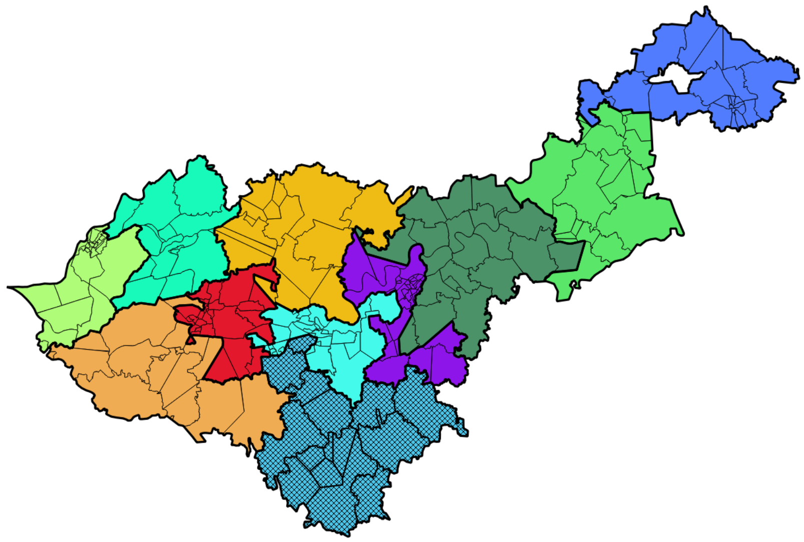


Figure 19: Tactical district design obtained with model AV–AW on instance 1 with a heterogeneous fleet of vehicles

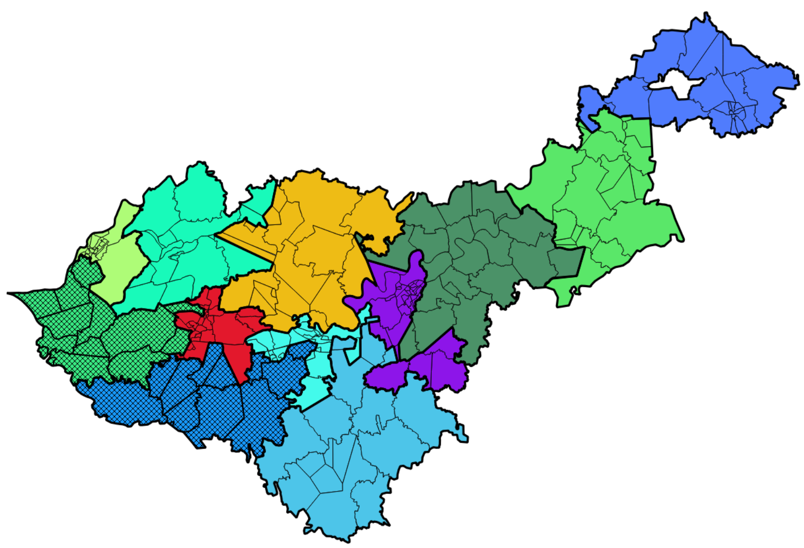


Figure 20: Tactical district design obtained with model A/IV–AW on instance 1 with a heterogeneous fleet of vehicles

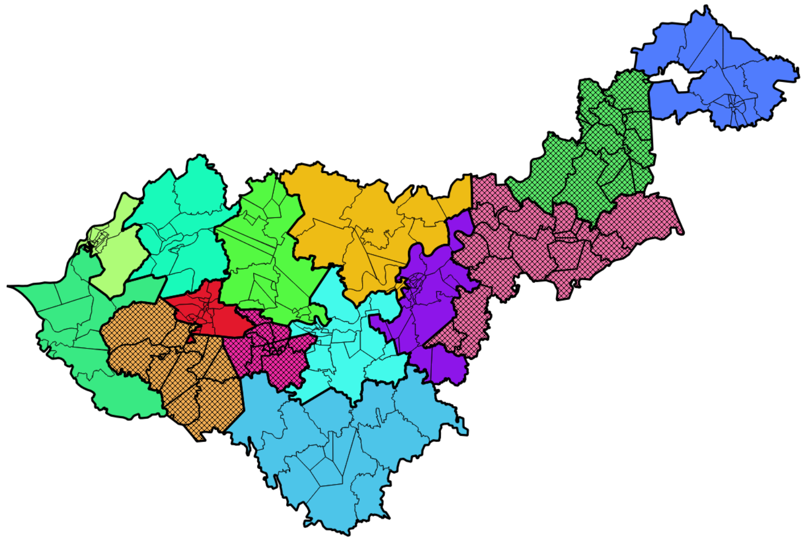


Figure 21: Tactical district design obtained with model IV–A/IW on instance 1 with a heterogeneous fleet of vehicles

E Visualization of Operational Reassignments

In addition to Figure 11 from Section 7.7, Figures 22–24 illustrate the operational reassignments obtained for $\omega \in \{5, 15, 20\}$. It can be seen from the figures that not all basic areas that are reassigned for small values of ω are also reassigned for greater values of ω. Obviously, increasing values of ω permit additional combinations of reassignments that are, at least in parts, more attractive than the reassignments that are feasible for smaller values of ω.

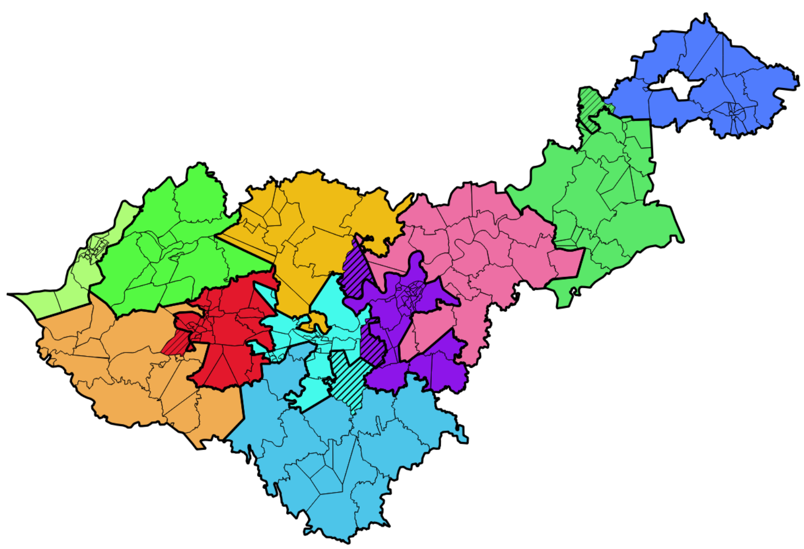


Figure 22: Solution obtained using model A/IV–AW on instance 1 after operational reassignment with $\omega = 5$

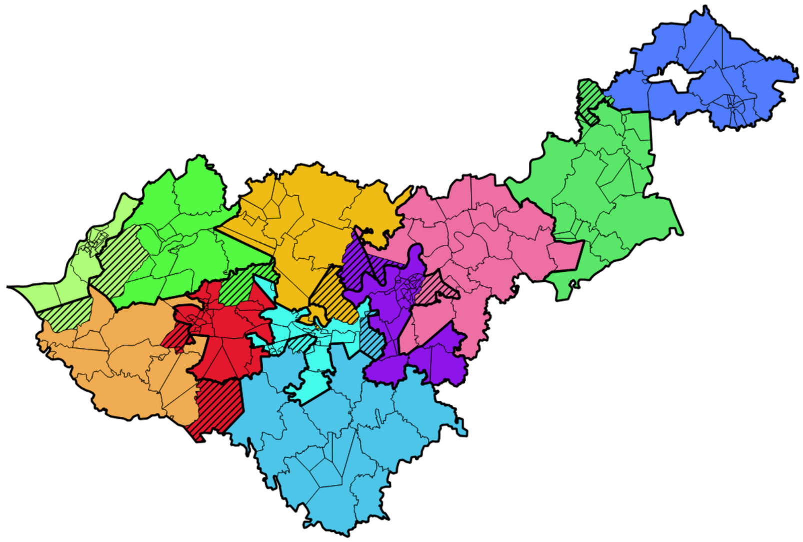


Figure 23: Solution obtained using model A/IV–AW on instance 1 after operational reassignment with $\omega = 15$

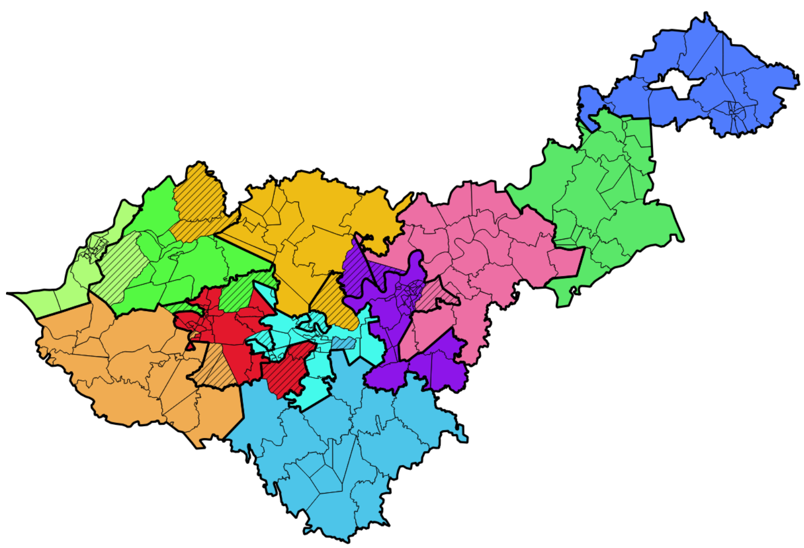


Figure 24: Solution obtained using model A/IV–AW on instance 1 after operational reassignment with $\omega = 20$